

Titre: Caractérisation des lieux d'activités à partir de données de cartes à puce de transport collectif
Title:

Auteur: Clément Piriou
Author:

Date: 2008

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Piriou, C. (2008). Caractérisation des lieux d'activités à partir de données de cartes à puce de transport collectif [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. <https://publications.polymtl.ca/8269/>
Citation:

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/8269/>
PolyPublie URL:

Directeurs de recherche:
Advisors:

Programme: Non spécifié
Program:

UNIVERSITÉ DE MONTRÉAL

CARACTÉRISATION DES LIEUX D'ACTIVITÉS
À PARTIR DE DONNÉES DE CARTES À PUCE
DE TRANSPORT COLLECTIF

CLÉMENT PIRIOU
DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE INDUSTRIEL)
AVRIL 2008



Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-41574-0

Our file Notre référence

ISBN: 978-0-494-41574-0

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

CARACTÉRISATION DES LIEUX D'ACTIVITÉS
À PARTIR DE DONNÉES DE CARTES À PUCE
DE TRANSPORT COLLECTIF

présenté par : PIRIOU Clément

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. ADJENGUE Luc, Ph.D., président

M. AGARD Bruno, Doctorat, membre et directeur de recherche

M. TRÉPANIER Martin, ing. Ph.D., membre et codirecteur de recherche

Mme MORENCY Catherine, ing. Ph.D., membre

REMERCIEMENTS

Je tiens particulièrement à remercier mes directeurs et codirecteurs de recherche, M. Bruno Agard et M. Martin Trépanier, pour m'avoir permis de faire cette maîtrise mais aussi pour leurs conseils techniques et méthodologiques, leur disponibilité et leur bonne humeur. Je tiens aussi à remercier la STO pour les données qu'elle a mis à ma disposition.

RÉSUMÉ

L'utilisation des cartes à puce comme titre de transport se développe depuis la fin des années 90. Les buts premiers de ces systèmes sont la perception de frais dus à l'utilisation du réseau et la vérification instantanée de la validité des titres de transport. Ces systèmes produisent un très grand nombre de données qui peuvent aussi servir à analyser le réseau de transport. Effectivement, grâce aux bases de données issues de tels systèmes, on dispose d'une vue globale et continue des comportements des usagers. Ces comportements sont évidemment conditionnés par la géographie du territoire desservi. Par exemple, les usagers utilisent surtout le transport collectif pour faire des trajets domicile-travail. C'est pourquoi le centre-ville attire la plupart des usagers du transport collectif pendant l'heure de pointe du matin. Ce sont ces variations spatio-temporelles qui vont nous intéresser.

Nous allons étudier l'influence des principaux lieux d'activités sur un réseau de transport en commun à partir des données de cartes à puce de la Société des Transports de l'Outaouais. Autrement dit, nous allons voir quels sont les comportements habituels des usagers qui montent aux arrêts situés à proximité des générateurs de déplacements principaux. A noter que les données sont complètement anonymes et respectent la stricte confidentialité des utilisateurs.

Comme la base de données où sont enregistrées toutes les validations des titres de transport comporte beaucoup de transactions (plus de 800 000 par mois), nous allons utiliser des techniques de data mining pour en extraire des connaissances compréhensibles et utilisables sur le réseau de transport en commun.

Dans une première partie, nous allons faire une revue de littérature traitant principalement des modélisations et des analyses de la variabilité des comportements en

transport, des systèmes de cartes à puce et leurs intérêts pour le transport en commun et enfin du data mining et plus particulièrement son application sur des bases de données géographiques.

Dans une deuxième partie, nous allons présenter les caractéristiques du système de transport en commun de la STO et expliquer la méthodologie utilisée pour en caractériser les lieux d'activités. La méthode utilisée se segmente en deux phases principales:

- D'une part, nous allons étudier les comportements des arrêts de bus en segmentant la population d'arrêts en un petit nombre de groupes aux comportements homogènes grâce à des algorithmes de data mining. Un des principaux problèmes rencontrés lors de cette phase est le choix d'un nombre de groupes adéquat. Après avoir discuté et trouvé un nombre de groupes adéquat, nous allons étudier la variabilité des arrêts entre les groupes pour faire diminuer le bruit dans les données en identifiant les arrêts les plus variables.
- D'autre part, nous allons créer un système d'information géographique (SIG) couvrant le territoire du réseau de transport en commun de la STO. Ce SIG localisera les principaux lieux d'activités de l'Outaouais et les arrêts de bus de la STO. Puis, nous définirons des relations de proximité entre les différents éléments du SIG pour enfin parvenir à une caractérisation des lieux d'activités.

La fin de la deuxième partie présente les bases de données utilisées pour mener les études, décrit le montage du SIG et explicite les tendances principales du réseau de transport en commun (profil des montants selon le jour de la semaine, Origine-Destination, etc.)

La troisième partie du mémoire s'attache à appliquer la méthodologie définie et à en extraire les résultats. Ainsi, nous allons segmenter la population d'arrêts de bus en neuf groupes de comportements différents. L'étude de la variabilité de l'appartenance à ces

groupes nous montre que les groupes sont constitués d'un noyau d'arrêts qui sont très souvent dans ce groupe et d'un nuage d'arrêts dont le comportement varie beaucoup. En identifiant les noyaux des groupes, nous serons en mesure de diminuer le bruit dans les données. L'étude de la variabilité selon le jour de la semaine nous confirme les différences d'utilisation du réseau de transport collectif entre les fins de semaine et le reste de la semaine. Le week-end, les usagers ont des comportements beaucoup plus irréguliers. En analysant la répartition des catégories de cartes (adulte, aîné, étudiant), nous verrons que les étudiants sont majoritaires aux arrêts qui sont utilisés surtout entre 9h et 16h et qu'ils ont tendance à faire des voyages plus courts. L'analyse des données confirme aussi que les usagers qui habitent loin du centre ville utilisent les transports plus tôt le matin et pour faire des distances plus grandes en général. Et enfin, la caractérisation des lieux d'activités montre que les stationnements incitatifs engendrent surtout des montants entre 7h et 8h du matin, le centre ville surtout entre 16h et 17h, les hôpitaux et les centres commerciaux engendrent une utilisation assez régulière du réseau pendant la journée.

ABSTRACT

The use of smart-card in transit system has developed a lot since the end of the 90's. Their first goal is to collect fare automatically and to check instantly the ticket validity. These systems produce a huge amount of data that can be used to analyze a transit network. Actually, with the databases stemming from such systems, we can have a full and continuous picture of travelers' behaviors. These behaviors are conditioned by the geography of the area and the structure of the network. For instance, people use public transportation mostly for commuting. That's why city centers attract most users during the morning peak hours. These variations in space and time will be the core of this study.

We are going to analyze the influence of major activity locations on a public transportation network using smart-card data from the Outaouais Transit Company (Société des transports de l'Outaouais, STO). That is to say, we are going to see how users who board near a main trip generator usually behave. The volume of data in the transportation database is so huge (more than 800 000 data a month) that we are going to use data mining techniques to extract knowledge from them.

In the first part, we are going to present a literature review about variability analysis and models of transportation users' behaviors, smart-card systems and their interests and finally, data mining and its application on geospatial databases.

In a second part, we are going to present the STO public transportation network and to explain the methodology used to characterize activity places. The method is divided in two main phases:

- On one hand, we are going to study bus stops' behaviors by clustering the bus stop population in a small number of homogeneous groups with data mining algorithms. During this phase, one of the major issues we face is the determination of an appropriate number of groups. Thus, after having discussed

and identified a appropriate number of groups, we are going to study the stops variability within these groups in order to decrease the data noise.

- On the other hand, we are going to create a geographic information system (GIS) of the STO public transportation network. This GIS locates all major activity places in the Ottawa region and all the STO bus stops. Then, we are going to define proximity relations between GIS elements to, finally, find a characterization for the activity places.

The end of this second part is presenting the databases used to undertake the studies and is describing the main facts of the network (number of boardings per hour per day, origin-destination...).

The third part is applying the methodology and extracting the results. Actually, we are going to cluster our bus stops population in nine typical behavior groups. The study of the belonging group variability will show us that groups are composed of a core of non-variable stops and a cloud of very unpredictable elements. It will permit us to decrease the noise in the data. By analyzing the proportion of each type of users (Adult, elderly, student) in each group, we will see that there is a majority of students on stops mostly used between 9:00 am and 16:00 pm and that they tend to make shorter trips. This analyze also confirms that users living far from the city center use transportation earlier in the morning and they tend to do longer trips. And finally, the activity places characterization reveals us that car parks create boardings mostly between 7:00 and 8:00 in the morning, the city center between 16:00 and 17:00, hospitals and shopping centers create a regular use during all the day.

TABLE DES MATIÈRES

REMERCIEMENTS	IV
RÉSUMÉ.....	V
ABSTRACT	VIII
TABLE DES MATIÈRES	X
LISTE DES TABLEAUX.....	XIII
LISTE DES FIGURES	XV
INTRODUCTION.....	1
CHAPITRE 1. REVUE DE LITTÉRATURE	4
1.1. Variabilité des comportements en transport	4
1.1.1. Méthodes de collecte des données	4
1.1.2. Le concept de journée moyenne.....	6
1.1.3. Étude de la variation journalière	7
1.1.4. Influence du transport collectif sur l'utilisation du sol	9
1.1.5. Les principaux lieux d'activités	9
1.2. Les systèmes de paiement par carte à puce en transport en commun	11
1.2.1. Présentation	11
1.2.2. Intérêts/limites.....	12
1.2.3. Problème de confidentialité et de sécurité des données	13
1.3. Data mining	14
1.3.1. En général	14
1.3.2. Data mining géo-spatial	15
CHAPITRE 2. CONTEXTE ET MÉTHODOLOGIE	17
2.1. Le système de transport en commun de la STO.....	17

2.1.1.	Géographie de l’Outaouais.....	17
2.1.2.	Technologie.....	19
2.1.3.	Modèle objet du système de carte à puce de la STO.....	19
2.1.4.	Structure tarifaire de la STO	22
2.2.	Méthodologie	25
2.2.1.	Étude de la similarité des comportements des arrêts	26
2.2.2.	Étude des lieux d’activités	28
2.3.	Les bases de données.....	28
2.3.1.	Base de données désagrégée	29
2.3.2.	Base de données agrégée.....	30
2.4.	Montage du SIG du réseau de la STO.....	31
2.4.1.	Organisation du SIG.....	31
2.4.2.	La couche des arrêts de bus du réseau	32
2.4.3.	La couche des arrêt-ligne-direction.....	33
2.4.4.	La grille de regroupement des données.....	33
2.4.5.	Les générateurs de déplacements	35
2.5.	Tendances principales.....	38
2.5.1.	Le profil moyen des montants.....	38
2.5.2.	Nombre d’arrêt-ligne-direction utilisés par jour	40
2.5.3.	Origine-destination.....	41
2.5.4.	Distance parcourue lors d’une section de déplacement	42
2.5.5.	Les classes d’utilisateurs	43
CHAPITRE 3.	EXPÉRIMENTATION ET RÉSULTATS	47
3.1.	Segmentation en groupes d’arrêt-ligne-direction homogènes.....	47
3.1.1.	Préparation des données.....	47
3.1.2.	Fouille de données et sélection d’un nombre adéquat de clusters	48
3.1.3.	Identification des neuf groupes d’arrêt-ligne-direction.....	59

3.1.4.	Comparaison des segmentations en trois et neuf groupes.....	62
3.2.	Étude de la variabilité au cours des 4 mois d'observation	64
3.2.1.	Variabilité des groupes d'arrêt-ligne-direction	64
3.2.2.	Étude de la variabilité le long des ligne-direction.....	75
3.2.3.	Variabilité des arrêts dans les groupes d'arrêt-ligne-direction	78
3.2.4.	Influence du jour de la semaine sur les groupes d'arrêt-ligne-direction..	80
3.3.	Étude des comportements des usagers dans les groupes d'arrêt-ligne- direction	84
3.3.1.	Étude des types d'utilisateurs.....	85
3.3.2.	Origine	86
3.3.3.	Les distances de déplacement	88
3.3.4.	Étude des groupes à majorité étudiante.....	90
3.4.	Influence spatio-temporelle des lieux d'activités sur un réseau de transport en commun.....	95
3.4.1.	Mise en place des relations de proximité	95
3.4.2.	Études des usagers aux les arrêts proches des lieux d'activités	97
3.4.3.	Caractérisation des lieux d'activités.....	102
CONCLUSION.....		110
BIBLIOGRAPHIE.....		114

LISTE DES TABLEAUX

Tableau 2.1 Grille tarifaire de la STO (2007)	23
Tableau 2.2. Base de données agrégée.....	31
Tableau 2.3. Répartition des usagers dans les classes.....	44
Tableau 3.1. Proportion de montants par heure, par arrêt-ligne-direction et par jour.....	47
Tableau 3.2. Groupe d'arrêt-ligne-direction (en ordre décroissant du nombre de montants).....	49
Tableau 3.3. Nouveau découpage temporel d'une journée avec un regroupement des heures du milieu de la journée	51
Tableau 3.4. Groupe d'arrêt-ligne-direction (en ordre décroissant du nombre de montants).....	53
Tableau 3.5. Nouveau découpage temporel d'une journée avec un regroupement des heures du milieu de la journée et du soir.....	55
Tableau 3.6. Groupe d'arrêt-ligne-direction (en ordre décroissant du nombre de montants).....	56
Tableau 3.7. Groupe d'arrêt-ligne-direction (en ordre décroissant du nombre de montants).....	58
Tableau 3.8. Groupe d'arrêt-ligne-direction (en ordre décroissant du nombre de montants).....	59
Tableau 3.9. Matrice d'interpénétration des segmentations en 3 et 9 groupes	62
Tableau 3.10. Variabilité de l'appartenance à un groupe des arrêt- ligne-direction	69
Tableau 3.11. Nombre d'éléments dans le noyau de chaque groupe	71
Tableau 3.12. Variabilité de l'appartenance à un groupe des arrêts d'une ligne-direction	76
Tableau 3.13. Relation de proximité des arrêt-ligne-direction avec les générateurs de déplacements	96
Tableau 3.14. Groupes prépondérants et générateurs de déplacements.....	102

Tableau 3.15. Pourcentage des différents groupes prépondérants proches des générateurs de déplacements	103
Tableau 3.16. Groupes prépondérants du lundi et générateurs de déplacements.....	105
Tableau 3.17. Pourcentage des différents groupes prépondérants du lundi proches des générateurs de déplacements.....	106

LISTE DES FIGURES

Figure 2.1. Quartiers desservis par la STO	18
Figure 2.2. Centre ville et distances caractéristiques de l’Outaouais.....	18
Figure 2.3. Modèle objet du système de transport de la STO	20
Figure 2.4. Représentation schématique de la méthodologie utilisée	25
Figure 2.5. Modèle relationnel simplifié.....	29
Figure 2.6. Répartition géographique des arrêts de la STO	32
Figure 2.7. Grille de regroupement des données.....	34
Figure 2.8. Arrêts par case de la grille	35
Figure 2.9. Les nœuds du réseau.....	37
Figure 2.10. Les principaux générateurs de déplacements	38
Figure 2.11. Profil des montants par heure et par jour de la semaine (somme sur toute la période d’observation)	39
Figure 2.12. Nombre d’arrêt-ligne-direction utilisés par jour.....	40
Figure 2.13. Origines (rouge) et destinations (bleu) des montants entre 8h et 9h dans chaque case de la grille pendant le mois de septembre 2005	41
Figure 2.14. Origines (rouge) et destinations (bleu) des montants entre 16h et 17h dans chaque case de la grille pendant le mois de septembre 2005	42
Figure 2.15. Distance moyenne (en mètre) parcourue au cours d’une section de déplacement au départ des différentes zones	43
Figure 2.16. Profil des montants par classe pour les jeudis de septembre 2005	44
Figure 2.17. Proportion des classes d’usagers par heure pour tous les jeudis de septembre 2005.....	45
Figure 3.1. Dendrogramme fait avec 30 sous-groupes.	48
Figure 3.2. Profil de la distribution des montants en fonction de l’heure de la journée pour 21 clusters	50
Figure 3.3. Dendrogramme fait avec 30 sous-groupes	52

Figure 3.4. Profil de la distribution des montants en fonction de la période de la journée pour 14 clusters	54
Figure 3.5. Dendrogramme fait avec 30 sous-groupes.	55
Figure 3.6. Profil de la distribution des montants en fonction de la période de la journée pour 3 clusters	56
Figure 3.7. Profil de la distribution des montants en fonction de la période de la journée pour les 9 clusters identifiés.....	60
Figure 3.8. Pourcentage de montants par groupe et par heure de la journée	61
Figure 3.9. Variabilité de l'appartenance des arrêt-ligne-direction aux différents groupes par semaine pendant la période d'observation.	64
Figure 3.10. Variabilité de l'appartenance des arrêt-ligne-direction	66
Figure 3.11. Variabilité de l'appartenance des arrêt-ligne-direction au groupe sur tous les mardis.....	67
Figure 3.12. Variabilité de l'appartenance des arrêt-ligne-direction sur tous les samedis	68
Figure 3.13. Variabilité de l'appartenance des arrêt-ligne-direction à un groupe prépondérant.....	70
Figure 3.14. Structure « atomique » d'un groupe avec son noyau et son nuage variable	72
Figure 3.15. Variabilité de l'appartenance des arrêt-ligne-direction à un groupe prépondérant selon le jour de la semaine	74
Figure 3.16. Variabilité de l'appartenance des ligne-direction à un groupe prépondérant.	77
Figure 3.17. Variabilité de l'appartenance des ligne-direction à un groupe prépondérant selon le jour de la semaine.	78
Figure 3.18. Variabilité de l'appartenance des arrêts à un groupe prépondérant selon le jour de la semaine.	79
Figure 3.19. Nombre d'arrêt-ligne-direction utilisés par jour de la semaine et par groupe	80

Figure 3.20. Nombre de montants par groupe (autre que régulier) et par jour de la semaine.....	81
Figure 3.21. Nombre de montants pour le groupe régulier par jour de la semaine.....	81
Figure 3.22. Contribution à chaque jour de chaque groupe d'arrêt-ligne-direction.....	82
Figure 3.23. Groupes subissant une baisse relative du nombre d'arrêt-ligne-direction le week-end	83
Figure 3.24. Groupes subissant une augmentation relative du nombre d'arrêt-ligne-direction le week-end	84
Figure 3.25. Proportion des types d'utilisateurs dans les groupes d'arrêt-ligne.....	85
Figure 3.26. Représentation géographique de la proportion de montants dans les principaux groupes d'arrêt-ligne-direction pour le mois de septembre 2005	87
Figure 3.27. Représentation géographique de la proportion d'arrêt-ligne-direction-jour des groupes par secteur	87
Figure 3.28. Pourcentage cumulé de la distribution des distances parcourues par déplacement dans chaque groupe d'arrêt-ligne-direction	89
Figure 3.29. Nombre d'utilisateurs par zone d'origine (rouge) et de destination (bleu) du groupe pic 8h pour le mois de septembre 2005.....	91
Figure 3.30. Nombre d'utilisateurs par zone d'origine (rouge) et de destination (bleu) du groupe pic Milieu pour le mois de septembre 2005	92
Figure 3.31. Nombre d'utilisateurs par zone d'origine (rouge) et de destination (bleu) du groupe pic 15h pour le mois de septembre 2005.....	93
Figure 3.32. Proximité des générateurs de déplacements et des arrêts	97
Figure 3.33. Profil des montants aux arrêts proches des écoles par heure et par jour de la semaine.....	98
Figure 3.34. Proportion des classes d'utilisateurs montants aux arrêts proches des écoles pour tous les jeudis de septembre 2005 par heure de la journée.....	99
Figure 3.35. Profil des montants aux arrêts proches des centres d'achat par heure et par jour de la semaine	100

Figure 3.36. Proportion des classes d'usagers montants aux arrêts proches des centres commerciaux pour tous les jeudis de septembre 2005 par heure de la journée.....	101
Figure 3.37. Variabilité de la proportion des arrêt-ligne-direction proches des écoles dans les groupes prépondérants pour les différents jours de la semaine...	107
Figure 3.38. Variabilité de la proportion des arrêt-ligne-direction proches des centres commerciaux dans les groupes prépondérants pour les différents jours de la semaine.....	108

INTRODUCTION

La Société des Transports de l'Outaouais (STO) est une société de transport en commun de taille moyenne qui exploite environ 250 bus sur la région de l'Outaouais au Québec. Actuellement, elle emploie 605 personnes dont 357 chauffeurs. Elle est une des premières autorités de transport au Canada à avoir mis en place un système de perception par carte à puce, en 2001.

Un tel système récolte chaque jour une grande quantité de données sur les déplacements effectués par les usagers sur le territoire desservi. Même si le but principal de ce système est la perception de frais inhérents à l'utilisation du réseau, on peut, grâce à ces données, modéliser et comprendre les comportements et les habitudes des usagers du réseau. Ceci nous permet d'être en mesure de les desservir le mieux possible. Ces habitudes de voyages sont évidemment influencées par la géographie du territoire de la société de transport et plus particulièrement par l'emplacement des lieux d'activité. En effet, certains arrêts se trouvent dans des zones plus utilisées par les usagers car plus proches de certains générateurs de déplacements comme le centre ville, qui attire les travailleurs, ou encore les écoles où se rendent les étudiants.

Objectifs

L'objectif de ce mémoire de maîtrise est d'étudier les potentialités des données de carte à puce pour caractériser l'influence des lieux d'activités sur l'utilisation d'un réseau de transport en commun. Pour ce faire, les données des déplacements des utilisateurs de cartes à puce du réseau de bus de l'Outaouais seront expliquées, traitées et analysées. Un sous-objectif de cette étude est d'étudier la répétitivité spatio-temporelle des déplacements des usagers d'un système de transport en commun.

L'étude conjointe du territoire desservi par le réseau de transport en commun et des données de cartes à puce va nous permettre d'extraire des comportements spatio-temporels typiques d'utilisation du bus, et par la suite de les étudier en détails sous diverses perspectives comme par exemple du point de vue des types d'abonnements des usagers ou des lieux d'activités.

Contenu

Ce mémoire se divise en trois parties principales. Dans la première partie, nous ferons une revue de littérature dans des domaines pertinents pour l'analyse. Nous examinerons les domaines suivants :

- les différentes sources de données pour caractériser les réseaux de transport en commun et la demande de transport.
- les modélisations des variations des comportements des usagers du transport en commun et l'influence de l'utilisation du sol et des lieux d'activités sur ces variations
- les cartes à puce et leur contribution à la compréhension d'un réseau de transport en commun
- les techniques d'exploration de données (data mining) et plus précisément appliquées à la compréhension d'un territoire (data mining géo spatial)

La deuxième partie présentera, tout d'abord, le contexte particulier du transport collectif en Outaouais. Puis, elle montrera la méthodologie développée pour caractériser les lieux d'activités à partir de données de cartes à puce issues du réseau de transport collectif. Cette méthodologie se divise en trois grandes étapes. La première est l'étude des arrêts du réseau, la deuxième étape est la création d'un SIG et enfin la dernière étape est la mise en relation des deux premières parties pour caractériser l'influence des lieux d'activités sur l'utilisation du réseau. Ensuite, la deuxième partie présentera les bases de données issues du système de paiement par carte à puce. Puis, elle s'attachera à décrire

le montage du SIG de l'Outaouais. Et, finalement, elle explicitera les tendances principales du réseau de transport de la STO.

Et enfin, la troisième partie appliquera cette méthodologie et explicitera les principaux résultats. On appliquera, tout d'abord, des techniques de data mining pour segmenter les arrêts en un petit nombre de groupes homogènes, on déterminera ici un nombre adéquat d'agrégats. Ensuite, on se propose d'identifier clairement chaque groupe. Dans un troisième temps, on s'attachera à étudier la variabilité de ces groupes d'arrêts selon divers facteurs, notamment pour corriger le bruit dans les données. Cela va nous permettre d'étudier la répétitivité spatio-temporelle des comportements de déplacements. Puis, nous allons analyser les comportements des usagers dans les différents groupes d'arrêts. Et finalement, la dernière section va mettre en place des relations de proximité entre les éléments du SIG de l'Outaouais. Ceci va nous permettre d'étudier l'influence des différents générateurs de déplacements sur les comportements des usagers (profils des montants et classe d'usagers). Pour terminer, l'étude conjointe des groupes d'arrêts créés grâce au data mining et de leur proximité aux principaux lieux d'activités caractérisera l'influence spatio-temporelle des générateurs de déplacements sur un réseau de transport en commun.

CHAPITRE 1. REVUE DE LITTÉRATURE

1.1. Variabilité des comportements en transport

Pour bien planifier les opérations de transport, il est très important de connaître les variations de la demande. En effet, les planificateurs de transport en commun font face au compromis suivant : offrir le meilleur niveau de service aux usagers avec les moyens les plus limités possibles. La compréhension des comportements des usagers permet donc de faire correspondre le plus possible l'offre en transport aux fluctuations de la demande.

1.1.1. Méthodes de collecte des données

Pour comprendre les comportements des usagers, il est nécessaire de récupérer des données significatives sur l'utilisation des transports. Or la grande majorité des systèmes de transports ne permettent pas d'enregistrer la façon dont les usagers s'en servent. Dès lors, il est nécessaire de mener des enquêtes pour obtenir ces données. D'après Chapleau (1995), au Québec, les grandes enquêtes téléphoniques Origine-Destination sont la source principale de données pour mener à bien la planification en transport en commun. Ces enquêtes sont réalisées périodiquement dans de nombreuses régions, par exemple à Montréal, la période est de cinq ans. Lors de ces enquêtes, on interroge un échantillon représentatif de la population sur ces déplacements au cours d'une journée particulière (habituellement la veille).

Ainsi, d'après le rapport sur la mobilité des personnes à Montréal qui fait suite à l'enquête Origine-Destination de 2003, à Montréal, environ 5% de la population, soit 70 000 foyers ou 165 000 personnes, a été interrogé pour l'enquête de 2003. Les questions portent sur les caractéristiques des déplacements de la journée précédente (origine, destination, heure, mode, motif, lignes utilisés...), sur les données socio-

démographiques du foyer interrogé (nombre de personnes du foyer, revenu, localisation, nombre de véhicules) et des personnes du foyer (âge, sexe, occupation, permis de conduire...). Un des intérêts principaux de ce type d'enquête est le grand nombre de données récoltées. Ceci permet de faire des statistiques sur les déplacements, de modéliser l'impact des projets de transport urbain et de suivre les évolutions socio-démographiques des habitudes de transport.

Trépanier et Chapleau (2001) expliquent qu'avec le développement des nouvelles technologies, on utilise maintenant un seul fichier de données où sont stockés les déplacements individuels totalement désagrégés. C'est-à-dire que le fichier de données est plat, chaque enregistrement représente un déplacement individuel avec toutes ses caractéristiques. Ceci permet l'élargissement des possibilités de traitement et de modélisation à partir des enquêtes Origine-Destination.

Les sociétés de transport en commun utilisent aussi d'autres méthodes pour recueillir de l'information comme les comptes à bord ou les comptes à l'arrêt : une personne est chargée de compter à un endroit du réseau et une heure donné le nombre de personne dans un bus. On peut donc déterminer les points de charge maximale du réseau de transport en commun et dimensionner le réseau en fonction. Mais ces méthodes sont beaucoup moins explicites que les enquêtes Origine-Destination en ce qui concerne les comportements de déplacements de usagers.

Pour les chercheurs Murakami, Wagner et Neumeiseter (2000), ces sondages où l'utilisateur donne beaucoup d'informations sont forcément faussés, car les usagers ne rapportent pas les trajets les plus courts et connaissent mal leur horaire exacte. C'est pourquoi ils encouragent le développement de méthodes de sondage où les usagers interviennent le moins possible, comme, par exemple, un système où des utilisateurs témoins ont en leur possession un dispositif muni d'un GPS qui enregistre tout seul les données spatio-temporelles. Les usagers n'ont plus alors qu'à remplir les raisons de

voyages. De telles méthodes sont moins demandantes en termes de temps de traitements humains et les données récoltées sont de bien meilleure qualité mais il est très difficile de faire ce type de sondage à grande échelle à cause des coûts de l'équipement requis, ce qui nuit à la représentativité statistique des données récoltées.

Néanmoins, certains systèmes permettent de récolter des données sur les usagers de manière plus ou moins directe :

- Certains bus sont équipés de GPS et d'un système de comptage automatique (montée ou montée/descente). Le système compte à chaque arrêt le nombre de montants et/ou de descendants et enregistre pour chaque opération le lieu et l'heure. Ceci permet par exemple de mesurer le retard d'un bus par rapport à l'horaire annoncé.
- Système de perception des frais par cartes à puce : des tels systèmes enregistrent toutes les validations des titres de transports pour contrôler l'accès du réseau. Certains systèmes enregistrent aussi d'autres informations comme les heures ou les lieux où ont lieu les transactions. Plus le niveau de détail du système d'information est grand, plus les analyses vont être fines et précises.

1.1.2. Le concept de journée moyenne

La planification dans les transports se fait traditionnellement à partir d'une journée type représentative comme par exemple une journée moyenne de semaine (Jones et Clarke, 1988). Par le passé, on éliminait les variations journalières soit en réalisant des études seulement sur un jour donné (par exemple le jeudi), soit en agrégeant les données des différents jours de la semaine pour obtenir un jour moyen tant au niveau des individus que des jours de la semaine. Jones et Clarke montrent aussi que, même si les patrons des comportements agrégés (niveau macroscopique) sont relativement stables, des variations dans les comportements individuels (niveau microscopique) existent et se compensent.

Donc, pour appréhender globalement les patrons, on doit être en mesure de suivre individuellement les usagers.

1.1.3. Étude de la variation journalière

En parallèle avec l'évolution des technologies, on a vu apparaître le besoin de suivre les variations journalières de la demande en transport, toujours dans le but d'offrir aux usagers le meilleur service possible. En effet, selon Gärling et Axhausen (2003), les patrons de déplacements sont essentiellement dus aux habitudes des usagers, car, évaluer les façons alternatives de voyager a un coût élevé et le gain associé est incertain. C'est pourquoi il est très intéressant de comprendre ces habitudes pour la modélisation et la planification des transports. Mais mesurer les habitudes de voyages nécessite de suivre les variations sur de longues périodes, sinon on pourrait rater certains patrons à long terme.

Ainsi, le sondage Mobidrive, effectué en 1999 à Karlsruhe (Axhausen et al., 2002), a suivi les comportements journaliers pendant 6 semaines et a récolté des nombreuses données socio-économiques et géographiques sur les voyageurs. Les études faites par Axhausen et al. (2002) ont montré les potentialités de tels sondages pour expliquer les rythmes quotidiens de vie.

Schlich et Axhausen (2003) introduisent des indicateurs pour mesurer la similarité des comportements à partir de ce sondage. Par exemple, ils observent les variations du nombre de voyages par jour, mesurent la répétition des voyages et concluent que globalement les comportements sont répétitifs (surtout du lundi au vendredi) mais que plus on cherche de précision dans les similitudes moins on a de ressemblance entre les comportements. Les auteurs suggèrent aussi qu'il faut observer les comportements sur au moins deux semaines si on veut mesurer la similarité.

Kitamura *et al.* (2006) ont examiné les variations spatio-temporelles (prisme d'espace-temps) des comportements d'un jour à l'autre grâce au sondage Mobidrive. Ils montrent que l'heure de départ du premier voyage le matin varie bien plus que le lieu de départ de ce premier voyage et qu'une grande partie des variations de l'heure de départ du matin est aléatoire.

Un réseau de transport en commun interagit avec l'environnement dans lequel il se trouve. En effet, les usagers du transport collectif se rendent d'un lieu A pour aller faire une activité à un lieu B. Ce sont les contraintes spatio-temporelles sur les activités qui entraînent des contraintes sur le système de transport. Certains lieux d'activités sont utilisés seulement par certaines catégories d'usagers à certaines périodes de la journée et donc le transport vers et depuis ces zones doit être planifié en fonction. Des chercheurs ont donc approché la modélisation et la planification d'un point de vue « activité ».

Ainsi, Kitamura (1988) explique que cette approche a permis d'améliorer la compréhension des comportements des usagers et leurs implications sur les réseaux de transport ainsi que les modèles de prévisions en transport. Jun, Konstadinos et Goulias (1997) analysent les patrons de voyages et d'activités. En segmentant la population des voyageurs en quatre groupes (selon les activités et les comportements de voyages), ils observent une stabilité apparente des activités et des comportements, une variation plus grande des activités que des comportements de voyages. Ce lien entre voyages et activités est encore plus fort pour un jour donné que d'un jour à l'autre et encore plus faible d'une année à l'autre.

Lockwood *et al.* (2004) analysent les différences dans le comportement des usagers entre le week-end et le reste de la semaine, et l'explique notamment par rapport aux générateurs de déplacements. La semaine, les usagers font surtout des trajets domicile-travail alors que le week-end, ils utilisent le transport en commun pour les loisirs ou pour aller magasiner.

Morency *et al.* (2006, 2007) ont utilisé les données de cartes à puce de la Société des Transport de l'Outaouais pour analyser les variations journalières des comportements des usagers. Ils ont notamment étudié la régularité des usagers dans leur utilisation quotidienne du transport en commun. Ils ont montré que la plupart des usagers avaient des comportements très réguliers.

1.1.4. Influence du transport collectif sur l'utilisation du sol

Dans la littérature, on trouve beaucoup d'études à long terme sur l'influence du transport en commun sur l'utilisation du sol (Workman et Brod, 1997, Hunt *et al.*, 1994 ou encore Knight et Trygg, 1977). En effet, du point de vue des plans d'urbanisme, il est très intéressant de voir l'influence du transport sur la répartition des commerces, des types d'habitations, etc. Cela permet de prévoir les réactions aux modifications et de diriger les plans d'urbanisme. Ainsi, Cervero et Landis (1996) ont examiné sur 20 ans les variations d'utilisation du sol aux alentours des stations de transport en commun dans la baie de San Francisco (BART : Bay Area Rapid Transportation). Badoe et Miller (2000) ont recensé des études sur le rapport entre utilisation du sol et transport en commun dans différentes villes. Ils ont montré que la plupart du temps les politiques de développement du transport n'ont pas eu les conséquences attendues sur l'utilisation du sol, souvent ce sont d'autres facteurs qui influencent plus profondément l'urbanisme.

1.1.5. Les principaux lieux d'activités

Afin de comprendre l'influence des activités sur le transport en commun, on doit recueillir des données pertinentes. On a vu que ces données sont souvent recueillies par des sondages. Ces sondages sont souvent composés de questions à choix multiples, notamment pour identifier les raisons de chaque déplacement.

Les chercheurs Doherty et Miller (2000) ont mis en place le sondage CHASE. Ce sondage demande à des foyers de fournir les informations sur leurs déplacements : où,

quand, comment (mode de transport), qui (membres du foyer) et pourquoi? On peut répondre à la question pourquoi par une liste prédéterminée d'activités. Les principales activités identifiées dans ce sondage sont (chacune ayant des sous-catégories):

- Travail
- Éducation
- Besoins primaires : restauration...
- Service : médecins, hôpital, dentistes, banque...
- Magasinage
- Loisirs
- Social
- Autres

À ces activités principales correspondent des lieux d'activités donnés qui sont les principaux générateurs de déplacement d'un réseau de transport.

Lee et McNally (2003) ont utilisé quasiment les mêmes catégories d'activités pour mener le sondage REACT! fait à Irvine, Californie entre Avril et Juin 2000 avec 72 répondants seulement, ce qui est vraiment peu pour pouvoir faire des statistiques.

De nombreux auteurs ont étudié l'influence de divers lieux d'activités sur le transport dans le but de créer des modèles de prévision de transport. Ainsi, Balmer *et al.* (2000) ont analysé l'influence du regroupement des écoles en Virginie-Occidentale sur les transports pour conclure que les campus où sont regroupés plusieurs types d'écoles induisent sur le réseau de transport quasiment le même comportement de transport qu'une école seule. Trépanier *et al.* (2003) ont utilisé les données des grandes enquêtes Origine-Destination de la région de Montréal. Ils ont suivi l'évolution du transport vers les hôpitaux entre 1987 et 1998 pour modéliser l'impact du déménagement des principaux hôpitaux sur l'achalandage du métro et du réseau routier. Beaucoup de chercheurs se sont intéressés à l'influence des commerces sur le transport. Ainsi, Datta *et al.* (1998) et Johnson et Hammond (2001) ont étudié les conséquences des différents types de stations services sur le trafic routier. Vivian (2006) a analysé des comptages de

l'heure de pointe de l'après-midi pour déterminer l'influence des centres commerciaux sur le trafic en fonction de la surface au sol des centres d'achats.

1.2. Les systèmes de paiement par carte à puce en transport en commun

1.2.1. Présentation

La technologie des cartes à puce n'est pas nouvelle. Le concept de carte en plastique avec une micro puce a d'abord été breveté en 1968 par les chercheurs allemands Jürgen Dethloff et Helmut Grötrup (Shelfer et Procaccino, 2002). Ce n'est qu'en 1977 que les groupes français Motorola et Bull fabriquent la première carte à puce. Les premières cartes à puce sont des cartes à puce à contact, elles sont surtout développées dans le domaine bancaire pour sécuriser les paiements. Puis, les cartes sont devenues "sans contact". C'est à partir de ce moment qu'elles ont commencé à intéresser les compagnies de transport, car elles permettent une validation rapide de l'information. Les premiers tests de ce type de cartes dans un réseau de transport ont eu lieu au début des années 90 (Blythe, 2004) mais ce n'est qu'à partir de la fin des années 90 que le système a été développé à grande échelle.

Le but principal des cartes à puce dans le transport en commun est de gérer la perception des frais liés à l'utilisation du réseau. Leur utilisation permet de faciliter et de sécuriser les opérations de comptabilité, de limiter l'usage frauduleux des titres de transports et de gérer les différents tarifs. Chaque fois qu'une carte est utilisée, la transaction effectuée est enregistrée dans un système d'information. Dans certains systèmes de carte à puce, les véhicules (les bus par exemple) sont équipés de GPS. Ceci permet de connaître pour chaque transaction le lieu où elle se fait. Blythe (1998) décrit en détail les dispositifs nécessaires pour un système de carte à puce (type de cartes, lecteurs, systèmes embarqués et centraux...). Les cartes à puce se complexifient de plus en plus et comme le souligne Bonneau (2002), elles doivent contenir de plus en plus d'information pour par exemple pouvoir être utilisées sur plusieurs opérateurs de transports. Dès lors, on peut se servir de ces données pour mener des analyses sur le réseau.

1.2.2. Intérêts/limites

L'implantation des systèmes de cartes à puce en transport en commun et donc l'étude des données associées sont relativement nouvelles. Par conséquent, peu d'auteurs ont pu mener des analyses précises avec ce type de données. Pour l'instant, on trouve surtout des articles sur les méthodologies d'analyses des données de cartes à puce et sur les grandes tendances que l'on peut mesurer.

Ainsi, selon Bagchi et White (2004), les apports principaux des données de cartes à puce sont les suivants :

- avoir accès à un grand volume de données de voyage des usagers
- pouvoir lier les transactions à un utilisateur particulier ou à une carte particulière mais ceci soulève des préoccupations sur la confidentialité des données
- pouvoir suivre sur une longue période les usagers
- connaître une assez grande partie des usagers d'une société de transport.

En effet, l'étude des données de carte à puce permet un suivi longitudinal sur une longue période. On peut donc détecter les profils de variabilité des comportements sur la semaine, sur un mois, sur une année ou sur plus de temps. On peut aussi souligner le coût relativement faible d'accès à toutes ces données, car elles sont déjà stockées dans un système d'information. Bagchi et White (2005) insistent sur l'importance de bien définir les objets et les concepts avec les données de cartes à puce. En effet, il faut souvent fixer des seuils de façon arbitraire. Par exemple, ils se demandent à partir de quelle durée un usager qui n'utilise pas sa carte à puce devient un «non-usager».

Hofmann et O'Mahony (2005) soulignent l'importance de mesurer l'impact des changements faits sur un réseau de transport collectif. Pour ce faire, ils mettent en place des indicateurs de performance qui permettent de suivre l'évolution de l'efficacité d'un réseau de transport en commun. Ils utilisent les données de cartes à puce du réseau de bus pour mesurer le niveau de service d'une société de transport. Ils mesurent les temps

d'attente lorsqu'un usager effectue une correspondance, plus le temps d'attente des usagers est faible, meilleur est le réseau.

D'après Bagchi et White (2004, 2005), les limites principales des connaissances extraites d'un système de cartes à puce sont les suivantes :

- pas d'informations exactes sur les points de départ et les destinations des déplacements. On peut connaître les arrêts mais pas les lieux (domiciles, travail...).
- dans beaucoup de cas, seuls des usagers «privilegiés» (abonnements...) utilisent des cartes à puce, certains usagers irréguliers peuvent toujours acheter des tickets papier à l'unité. Les données de carte à puce ont donc tendance à représenter un échantillon un peu plus régulier que la moyenne.

Pour avoir des informations plus complètes, les enquêtes menées auprès des usagers pour déterminer les données manquantes sont toujours nécessaires. Ces deux types d'enquêtes sont donc complémentaires. D'une part, les données de carte à puces permettent aussi, grâce à leur spectre plus large, de valider la représentativité des échantillons interrogés lors des enquêtes et d'autre part, les données d'enquête permettent de compléter les données de cartes à puces.

1.2.3. Problème de confidentialité et de sécurité des données

Les données de cartes à puce posent surtout un problème de confidentialité et de liberté individuelle. Effectivement, avec ces données, on peut suivre exactement les déplacements d'un usager donné. Clarke (2001) met en avant le fait que ces données sont aussi sensibles que les données de cartes de crédits ou que les bases de données policières.

La Commission Nationale de l'Informatique et de Liberté (CNIL, 2003) recommande la prudence quant à l'usage des données de cartes à puce de transport collectif et suggère de séparer autant que possible les informations personnelles (nom, adresse...) des

informations de voyage. Mais certains cas particulier, comme la gestion de la fraude, nécessitent un recoupement.

1.3. Data mining

1.3.1. En général

Du fait de l'augmentation continue de la taille de base de données, on a dû développer des outils spécifiques pour en extraire de façon plus ou moins automatique des connaissances. Ces procédés de fouille de données sont regroupés sous le terme data mining. Cette technique utilise des outils statistiques, de gestion de base de données et des heuristiques spécialement développés pour l'extraction de connaissances (Fayyad *et al.*, 1996).

D'après Westphal et Blaxton (1998), le data mining a trois fonctions principales :

- Segmentation : séparer une population en sous populations ayant des comportements similaires pour pouvoir étudier les différents patrons caractéristiques de la population.
- Classification : mettre des étiquettes prédéfinies sur des éléments en fonction de leurs attributs (par exemple: bon/mauvais, rouge/blanc/rosé). La classification inclut la prédiction.
- Description : visualiser directement les relations entre les attributs d'une population pour comprendre les valeurs de certains attributs (par exemple : détection de déclarations d'assurance frauduleuses)

Le data mining a notamment beaucoup été appliqué dans le domaine du marketing pour étudier les comportements des clients et ainsi cibler les actions à entreprendre (soigner ses meilleurs clients, ceux qui risquent de ne plus acheter nos produits...) (Berry et Linoff, 1997).

1.3.2. Data mining géo-spatial

Les systèmes d'information géographique (SIG) se sont aussi beaucoup complexifiés depuis quelques années. Maintenant, la quantité de données récoltées dans les bases de données géographiques est telle qu'on a dû développer des outils spécifiques de data mining pour en extraire des connaissances. Le data mining spatial est défini par Koperski *et al.* (1995) et comme l'extraction de connaissances implicites, de relations spatiales ou d'autres patrons qui ne sont pas explicitement stockés dans le système d'information géographique.

En effet, un système d'information géographique contient des données géographiques (villes, rivières, forêts...) mais aussi beaucoup de données sociodémographiques (nombre d'enfants par foyer, salaires moyens...) ou environnementales (météorologiques, concentration de polluants, trafic routier...) organisées en couches. Le data mining spatial va surtout s'attacher à définir des règles d'association entre ces différentes couches de données.

La principale difficulté rencontrée est de lier les éléments des différentes couches en définissant des opérateurs de proximité entre les objets (intersection, contenance, chevauchement, proximité, voisinage, direction...) (Freska, 1992), d'autant que les couches contiennent des éléments de natures différentes (points avec coordonnées x et y, lignes, polygones...). Les seuils utilisés pour définir ces relations de proximité sont arbitraires, ils varient donc d'une utilisation à l'autre.

Par exemple, Koperski *et al.* (1995) ont construit un modèle géographique de l'Amérique du Nord et ils ont trouvé que la plupart des villes de Colombie Britannique adjacentes à de l'eau sont proches de la frontière américaine, la confiance de cette loi étant de 92%. Les chercheurs Koperski *et al.* (1996) ont fait un bilan des différents algorithmes utilisés dans le domaine du data mining géospatial, tant les algorithmes de

segmentation que les algorithmes de détermination de règles d'association. Weeks (2003) applique ce type de fouille de données à l'étude de la fertilité en Égypte, il crée un SIG de l'Égypte et dresse, selon des critères sociogéographiques, un bilan de la fertilité en Égypte.

Après avoir développer les apports de la littérature scientifique dans les domaines:

- les sources de données en transport en commun
- les modèles de variation des comportements des usagers et l'utilisation du sol par rapport au développement des transports
- les systèmes de paiement par cartes à puce et leurs contributions pour les analyses en transports en commun
- les techniques d'exploration de données dans les SIG (datamining géo-spatial)

Nous avons constaté que peu d'études des données de cartes à puce avaient été menées dans la littérature. La partie suivante va maintenant s'attacher à décrire le contexte particulier du transport en commun en Outaouais et à expliquer la méthodologie mise en place pour analyser les données issues du système de paiement par cartes à puce. L'objectif de ces analyses est de montrer l'influence des principaux lieux d'activités sur l'utilisation du réseau de transport en commun.

CHAPITRE 2. CONTEXTE ET MÉTHODOLOGIE

2.1. Le système de transport en commun de la STO

2.1.1. Géographie de l’Outaouais

La ville de Gatineau est située à la frontière entre le Québec et l’Ontario. Elle est issue de la fusion en 2002 des villes d’Aylmer, de Buckingham, de Gatineau, de Hull et de Masson-Angers. Gatineau est actuellement la quatrième ville la plus peuplée au Québec avec un bassin de population d’environ 240 000 habitants et l’agglomération qu’elle forme avec Ottawa en fait la quatrième la plus peuplée du Canada avec plus d’un million de personnes. Le territoire de service de la STO s’étend sur la grande ville de Gatineau, ainsi que sur les villes de Cantley, Chelsea et Ottawa. Ce territoire est limitrophe de la ville d’Ottawa; ce qui rend la situation géographique du réseau de la STO assez particulière. En effet, le centre ville d’Ottawa est un des points d’attraction les plus importants pour les usagers de la STO et seulement quelques lignes traversent la rivière des Outaouais pour le desservir (Ottawa possède son propre réseau de transport en commun: OC Transpo).

La carte suivante (Figure 2.1) montre les différents quartiers desservis par le réseau de transport en commun de la STO.

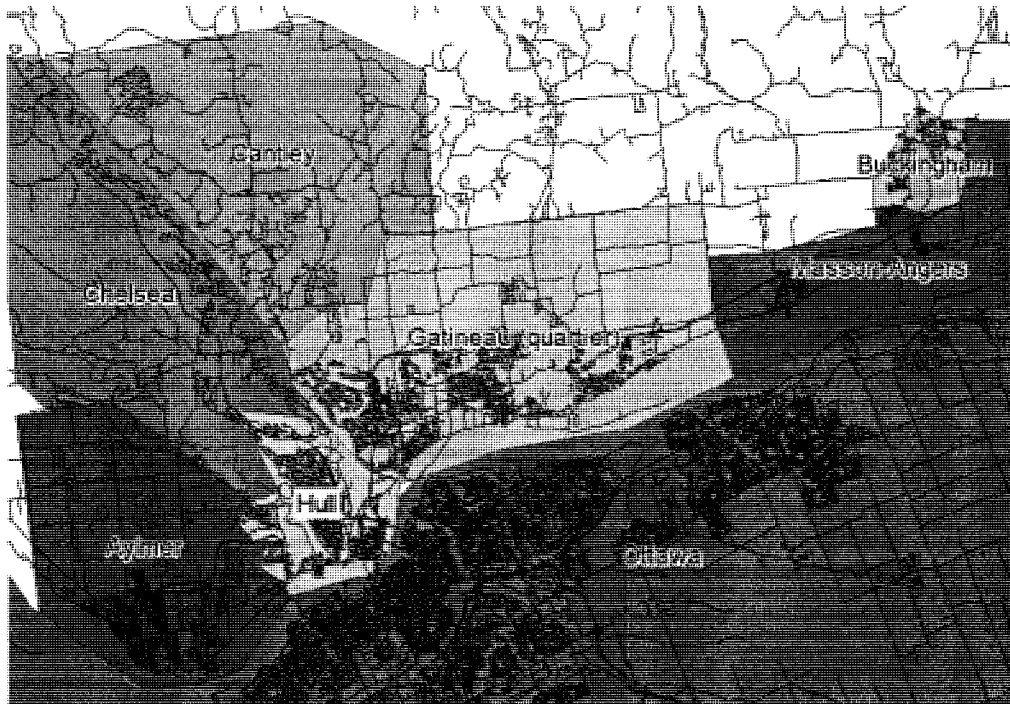


Figure 2.1. Quartiers desservis par la STO

La carte suivante (Figure 2.2) présente le générateur de déplacements principal : le centre ville de Gatineau et d'Ottawa et les distances caractéristiques de la ville.



Figure 2.2. Centre ville et distances caractéristiques de l'Ottawa

2.1.2. Technologie

La STO a mis en place un système paiement par carte à puce en 2001. Le système est composé d'un ensemble embarqué dans chaque bus qui possède une base de données (utilisateurs et parcours du bus dans la journée), enregistre les informations de chaque transaction (numéro de carte, date, heure) et autorise les transactions de montée. Il faut également souligner que ce système embarqué possède un GPS qui permet de déterminer l'arrêt où a lieu chaque transaction. Il faut noter qu'on n'enregistre pas la descente des usagers, on n'a donc aucune information directe sur les destinations. Les données récoltées par ce système sont transmises à la fin de chaque journée à un système central lors du retour du bus à son garage. La carte à puce utilisée par la STO est une technologie sans contact. A chaque montée dans un bus, l'utilisateur valide sa carte à puce en l'approchant du lecteur. Chaque utilisateur possède une carte nominative, avec une photo, que lui seul peut utiliser.

Pour des raisons de confidentialité et de liberté individuelle, la STO ne fournit pas de fichier où l'on peut faire la liaison entre les caractéristiques (nom, adresse,...) de l'utilisateur et son numéro de carte. En fait, ces données sont stockées dans deux bases de données différentes. Dès lors, on possède très peu d'informations sur les usagers.

2.1.3. Modèle objet du système de carte à puce de la STO

Le modèle objet du système de carte à puce de la STO a été développé pour mieux comprendre les relations entre les différents éléments du réseau de transport en commun. La méthode utilisée pour créer ce modèle est la méthode TOOM, Transportation Object-Oriented Modelling (Trépanier et Chapleau, 2001).

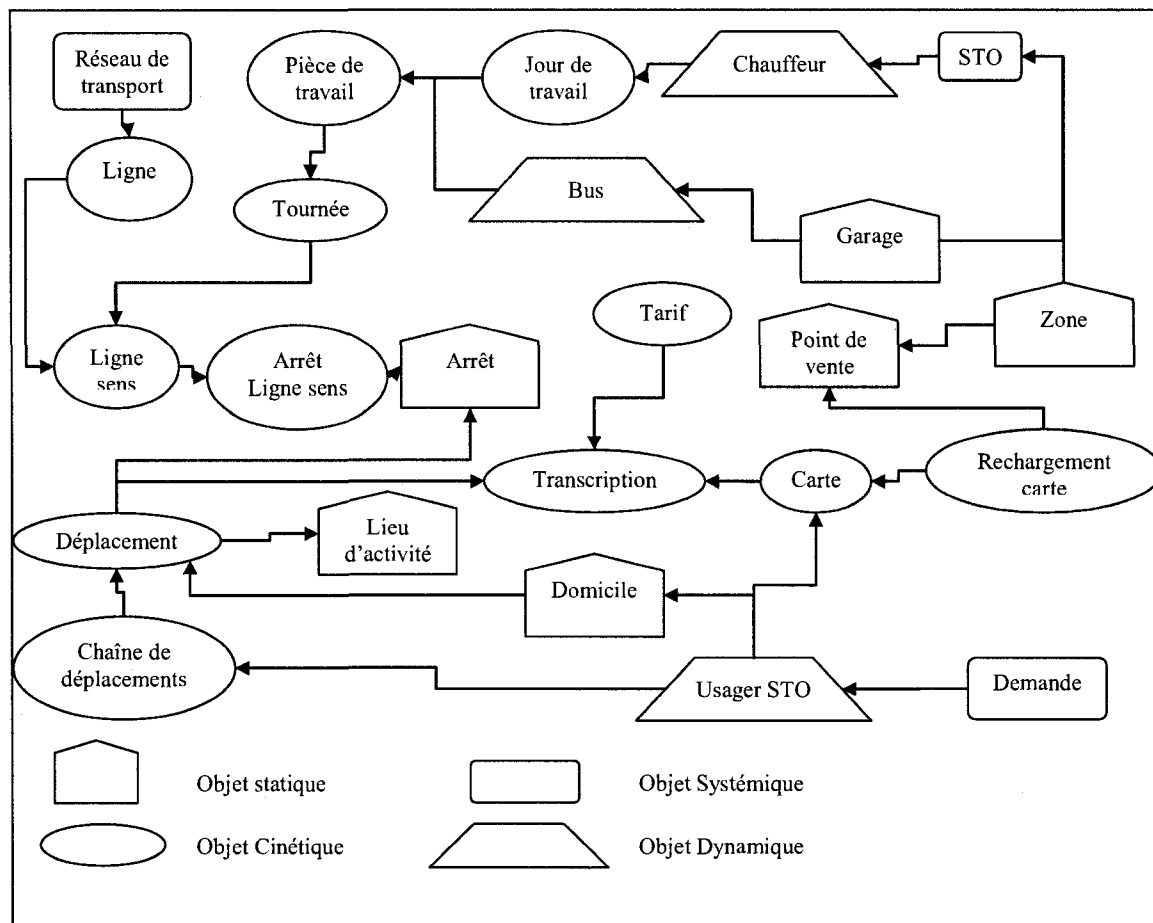


Figure 2.3. Modèle objet du système de transport de la STO

TOOM sépare les objets en trois méta-classes d'objets: les objets statiques, les objets cinétiques (ils décrivent les mouvements), les objets dynamiques (acteurs du transport) et les objets systémiques (réseaux, systèmes) (voir Figure 2.3).

La Figure 2.3 montre les objets du système de carte à puce de la STO. Pour simplifier, on peut regrouper les objets en 4 groupes principaux :

- Objets du réseau, éléments publics du réseau de transport en commun (ligne, arrêt, tournée...)
- Objets des opérations, éléments internes à la STO (chauffeur, bus, pièce de travail, garage)

- Objets administratifs, éléments propres du système de carte à puce
- Objets liés aux comportements des utilisateurs (demande, déplacement, chaîne de déplacements, domicile, lieu d'activité...)

Le modèle objet permet d'expliquer les relations entre les éléments du réseau et les données disponibles. Ainsi, pendant un mois par exemple, un utilisateur du service de transport en commun de la STO effectue une chaîne de déplacements. Une chaîne de déplacements est donc composée d'une série de plusieurs déplacements. Chaque déplacement s'effectue sur une ou plusieurs sections de ligne-direction du réseau qui sont associées à des tournées particulières. Les tournées sont elles-mêmes liées à des pièces de travail faites par un chauffeur particulier avec un bus particulier.

Pour notre étude, il est important de bien définir les notions suivantes :

- *Déplacement* : un usager effectue un déplacement lorsqu'il se rend d'un lieu d'activités à un autre, quelque soit le nombre de correspondances effectuées pour le trajet.
- *Section de déplacements* : une montée dans un bus, cela correspond à une validation de carte à puce. Un *déplacement* est donc composé d'une série de *sections de déplacement*.
- *Générateur de déplacements* : un générateur de déplacements est le lieu géographique qui correspond au but du déplacement d'un usager, il peut par exemple utiliser le bus pour aller à son bureau ou aller dans un centre d'achat.
- *Correspondances* : selon la définition de la STO, une montée qui se fait au plus 2 heures après une autre montée par le même usager (même carte) est une correspondance.
- *Ligne* : route empruntée par un bus depuis son terminus de départ jusqu'au terminus d'arrivée. Dans la base de données, elle est identifiée par un nombre variant entre 1 et 900. Une *ligne* est souvent composée de deux directions différentes (0 et 1 dans la base de données de la STO)

- *Arrêt* : un arrêt est un lieu où un bus peut s'arrêter pour embarquer ou déposer des usagers. Les arrêts sont repérés dans les bases de données de la STO par un nombre à quatre chiffres. Plusieurs *lignes* de bus différentes peuvent passer par un même *arrêt*.
- *Arrêt-ligne-direction* : un arrêt-ligne-direction est le lieu où peut s'arrêter un bus qui dessert une ligne donnée dans une direction donnée. Un *arrêt-ligne-direction* donné est emprunté par une seule *ligne*, dans une seule direction. Un *arrêt* peut donc correspondre à plusieurs *arrêt-ligne-direction*.

2.1.4. Structure tarifaire de la STO

La structure tarifaire de la STO est divisée en deux catégories principales : les abonnements (mensuels et semestriels) et les tickets papiers. Pour ces deux catégories, il existe des tarifs réguliers et des tarifs réduits. Les tickets ont une validité de 2 heures à partir de l'heure de départ du voyage en cours, période pendant laquelle l'utilisateur peut faire autant de correspondance qu'il le désire. Pour bénéficier d'une réduction, les usagers doivent se munir d'une carte de réduction. Cette carte est une carte à puce appelée carte « Passe-Partout PLUS », elle est nominative et coûte actuellement 8\$. Elle sert aussi de titre de transport rechargeable pour les abonnements.

Une autre dimension dans la structure tarifaire de la STO est le type de ligne sur lequel un usager monte :

- Régulier : lignes à intervalles réguliers qui composent la base du service
- Express : lignes opérant seulement aux heures de pointe
- Interzone : lignes qui empruntent l'autoroute pour lier rapidement le secteur Buckingham au centre ville (quartier Hull et Ottawa).

Le Tableau 2.1 présente les différents tarifs pour les lignes régulières, des suppléments s'appliquent sur les lignes Express et Interzone :

Tableau 2.1 Grille tarifaire de la STO (2007)

Tarif	Eligibilité	Titre de transport	Durée	Justificatif	Coût Lignes régulières
Régulier	Tous	Ticket Papier	2 heures	Aucun	2,50 \$
Régulier	Tous	Carte à puce	1 mois	Carte passe-partout PLUS	69,00 \$
Aîné	plus de 65ans	Ticket Papier	2 heures	Carte passe-partout PLUS Aîné	1,90 \$
Aîné	plus de 65ans	Carte à puce	1 mois	Carte passe-partout PLUS Aîné	31,00 \$
Enfant	moins de 5 ans	Aucun	2 heures	Aucun	Gratuit
Enfant	moins de 11 ans	Ticket Papier	2 heures	Aucun	1,90 \$
Étudiant	moins de 20 ans	Ticket Papier	2 heures	Carte passe-partout PLUS Étudiant	1,90 \$
Étudiant	moins de 20 ans	Carte à puce	1 mois	Carte passe-partout PLUS Étudiant	48,50 \$
Étudiant	plus de 20 ans	Carte à puce	4 mois	Carte passe-partout PLUS Étudiant	207,00 \$

Il existe aussi un abonnement spécial d'un jour : « le passe-temps » fait pour les touristes, il assure aux détenteurs un accès illimité aux réseaux de la STO et de OC Transpo. Ce billet n'est pas une carte à puce.

De plus, les usagers réguliers de la société de transport d'Ottawa, OC Transpo, qui possèdent un abonnement peuvent utiliser les réseaux de la STO sans supplément en présentant simplement leur laissez-passer valide (les tickets de OC Transpo ne sont pas valides pour la STO). Ces titres de transport ne sont pas des cartes à puce. Ces usagers qui arrivent d'Ottawa représentent environ 15% des usagers de la STO.

On doit noter que tous les usagers qui n'utilisent pas de carte à puce seront « invisibles » dans les bases de données et donc dans nos analyses. Autrement dit, les bases de données contiendront uniquement des données sur les usagers qui possèdent des cartes à puce. Actuellement, plus de 80% des utilisateurs de la STO possèdent une carte à puce. Au vu de ces données, l'échantillon d'usagers sur lequel

nous allons faire les analyses est suffisamment représentatif. Mais comme l'ont souligné Bagchi et White (2005) on a déjà un biais dans les données du fait que ce sont les usagers les plus réguliers qui souscrivent aux abonnements et qui ont donc une carte à puce. Les comportements observés par l'analyse des données de carte à puce vont donc être probablement plus lissés que les comportements réels.

On peut aussi constater que la donnée du type d'abonnement (adulte, aîné ou étudiant) est liée à la carte à puce d'un usager. C'est le seul renseignement socio-économique qui est disponible dans les bases de données fournies par la STO. Cette caractéristique va donc nous permettre de mener des analyses croisées plus poussées sur les utilisateurs.

2.2. Méthodologie

Le but de cette étude est d'analyser l'apport de l'analyse des données de carte à puce pour tirer des informations utilisables et compréhensibles sur l'utilisation géographique d'un réseau de transport en commun. On veut utiliser des données de carte à puce issues d'un réseau de transport collectif pour **comprendre comment les lieux d'activités influencent l'utilisation de ce réseau.**

Pour ce faire, nous allons, d'une part (voir Figure 2.4 : I), examiner la similarité des comportements des arrêts de bus d'un réseau et d'autre part (voir Figure 2.4 : II) étudier et expliciter les principaux générateurs de déplacements du territoire. Et enfin, nous regrouperons ces deux aspects pour caractériser l'influence des lieux d'activités sur le réseau.

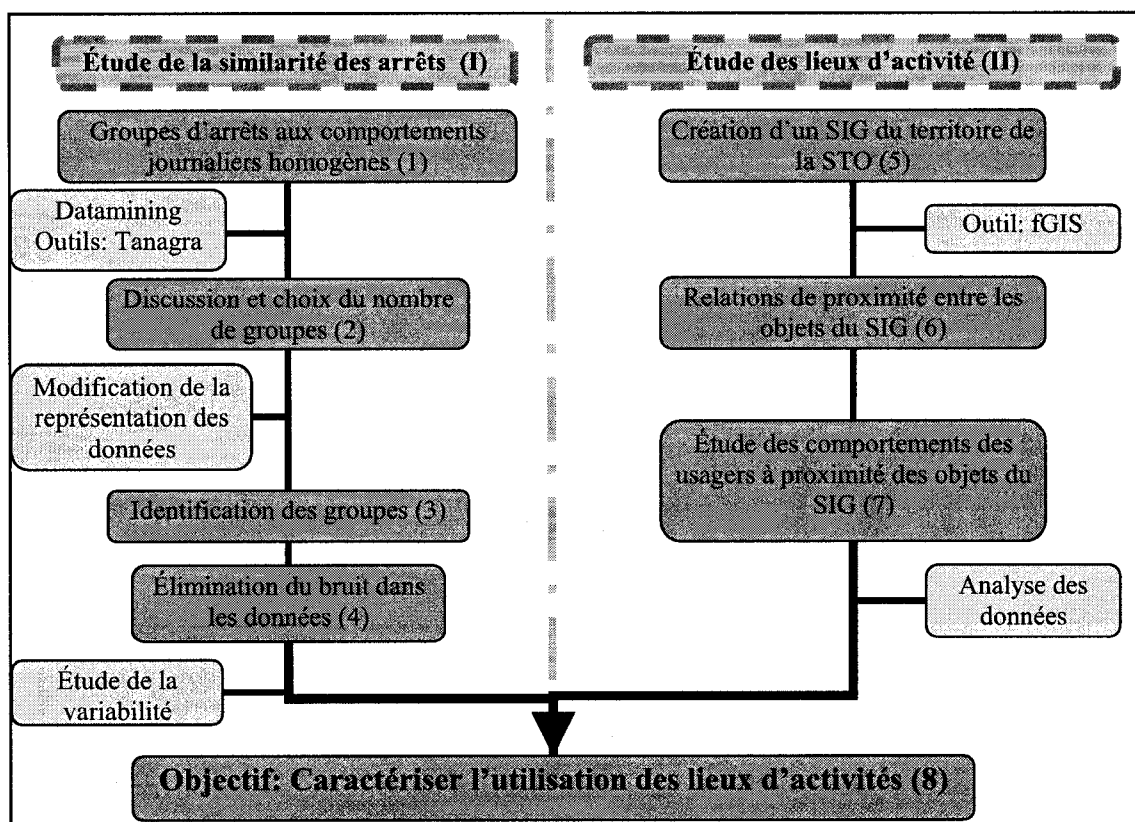


Figure 2.4. Représentation schématique de la méthodologie utilisée

Puisque les arrêts sont les composantes caractéristiques de la géographie du réseau car ils en sont les objets statiques, ce sont eux qui sont directement liés aux générateurs de déplacements. Autrement dit, nous allons nous demander si un lieu d'activités donné (par exemple une école) engendre sur les arrêts qui lui sont proches une façon caractéristique d'utiliser le transport collectif.

2.2.1. Étude de la similarité des comportements des arrêts

Dans une première partie (Figure 2.4: I), nous allons segmenter la population d'arrêt-ligne-direction en un petit nombre de groupes homogènes en termes de comportement journalier grâce à des techniques de data mining (voir Figure 2.4 : 1). La segmentation ne tient pas compte des coordonnées des arrêts car le but est de voir si des arrêts proches géographiquement vont naturellement se regrouper. Pour ce faire, on utilise le logiciel de data mining Tanagra (Rakotomalala, 2005). La méthodologie de segmentation est la suivante :

- On segmente la population étudiée en 30 sous-groupes avec l'algorithme k-means.
- On applique un algorithme de segmentation par agglomération hiérarchique sur ces 30 sous-groupes. Le logiciel de data mining nous fournit alors un dendrogramme et nous recommande un nombre de groupes « optimal ».
- On segmente la population de départ en ce nombre « optimal » de groupes avec un algorithme k-means.

Mais là intervient la question du nombre adéquat de groupes. Il existe des algorithmes qui permettent de déterminer un nombre adéquat de groupes d'un point de vue mathématique mais cela ne garantit pas que tous ces groupes soient significatifs d'un point de vue compréhension du réseau. Nous allons donc avancer pas à pas en commençant par le nombre de groupes suggéré par l'algorithme de segmentation par

agglomération hiérarchique. Puis, en travaillant essentiellement sur la représentation des données et en réappliquant la méthodologie de segmentation à ces nouvelles représentations, nous allons pouvoir déterminer un nombre adéquat de groupes qui permet de rendre plus compréhensible la base de données (voir Figure 2.4 : 2).

Ensuite, dans une deuxième partie, une fois les groupes d'étude déterminés, nous allons les caractériser pour comprendre en quoi les éléments de chaque groupe se ressemblent ou se différencient des éléments des autres groupes. Cette étape va aboutir à une identification de chaque groupe qui va permettre de bien comprendre de quels objets on parle (voir Figure 2.4 : 3).

Puis, la troisième partie analyse la variabilité de l'appartenance aux agrégats aux cours du temps (stabilité des comportements par semaine, par jour de la semaine...). On montrera ici que les groupes d'arrêts sont composés d'un noyau dur d'arrêts qui varient peu et d'un nuage très irrégulier. Ce noyau dur permettra de limiter le bruit dans les données en éliminant les arrêts les plus imprévisibles, qui changent souvent de type de comportement souvent du fait de leur faible utilisation (voir Figure 2.4 : 4). Ici, nous allons examiner si les comportements spatio-temporels des usagers sont répétitifs pendant la semaine, ou encore plus pour un jour de la semaine donné (par exemple pour les lundis) ou encore s'ils sont beaucoup plus variables le week-end.

Enfin, la quatrième partie présentera les caractéristiques des comportements des usagers dans les différents groupes d'arrêts. On analysera notamment les types d'utilisateurs qui montent sur des arrêts d'un groupe donné, les origines des déplacements, les distances des déplacements à partir des arrêts de chaque groupe.

2.2.2. Étude des lieux d'activités

La deuxième étape de la méthodologie consiste en l'étude de l'influence des principaux lieux d'activité sur le réseau de transport en commun (Figure 2.4 : II). Pour ce faire, nous allons, tout d'abord, construire un SIG du territoire de la STO (voir Figure 2.4 : 5) en identifiant, grâce aux sites web institutionnels de la région de l'Outaouais (Atlas de Gatineau, Commission scolaire des Portages-de-l'Outaouais, cégep de l'Outaouais) et en localisant, grâce aux outils Internet "Google Maps" ou encore wikimapia.org, un site de cartographie collaborative, les principaux lieux d'activité du territoire. Ce SIG va aussi contenir des couches relatives au réseau de transport en commun de la STO.

Ensuite, nous allons définir des relations de proximité (voisinage, inclusion...) entre les différents objets du SIG (voir Figure 2.4 : 6). Ceci va nous permettre d'étudier les comportements de usagers au voisinage de certains générateurs de déplacements (voir Figure 2.4 : 7). On étudiera notamment les profils des montants et les classes d'usagers qui utilisent le bus à proximité d'un lieu d'activité donné.

Enfin, on caractérisera les différents générateurs de déplacements en fonction de leur proximité aux groupes d'arrêts créés dans l'étape précédente (voir Figure 2.4 : 8). On pourra voir si les usagers d'un arrêt situé non loin d'un certain type de générateur de déplacements auront tendance à avoir un comportement type. Et enfin, on s'intéressera à la variabilité de ces caractérisations en fonction du jour de la semaine afin de mesurer si l'influence des générateurs de déplacements n'est pas la même le week-end que le reste de la semaine.

2.3. **Les bases de données**

La STO récolte des données sur environ 800 000 transactions par mois (montées). Au vu des temps de traitement avec le logiciel MS Access, traiter plus d'un mois de données est assez chronophage. Néanmoins, pour appréhender globalement les patrons dans ces

données, nous avons besoin de plus d'un mois d'historique. Par exemple, si on étudiait uniquement le mois de septembre 2005, les lundis fériés représenteraient 25% des lundis (un lundi sur 4 est férié). Les interprétations que l'on pourrait alors faire sur les jours de la semaine seraient donc très biaisées. C'est pour cela que les analyses de data mining seront effectuées sur des bases de données déjà agrégées, entre les mois de septembre et décembre 2005. On pourra ensuite croiser les résultats des fouilles de données avec la base de données désagrégée du mois de septembre 2005 pour faire des analyses plus fines. Ces analyses en détail se feront donc sur une sous-partie de l'ensemble global. La figure suivante (Figure 2.5) présente le modèle relationnel simplifié de la base de données.

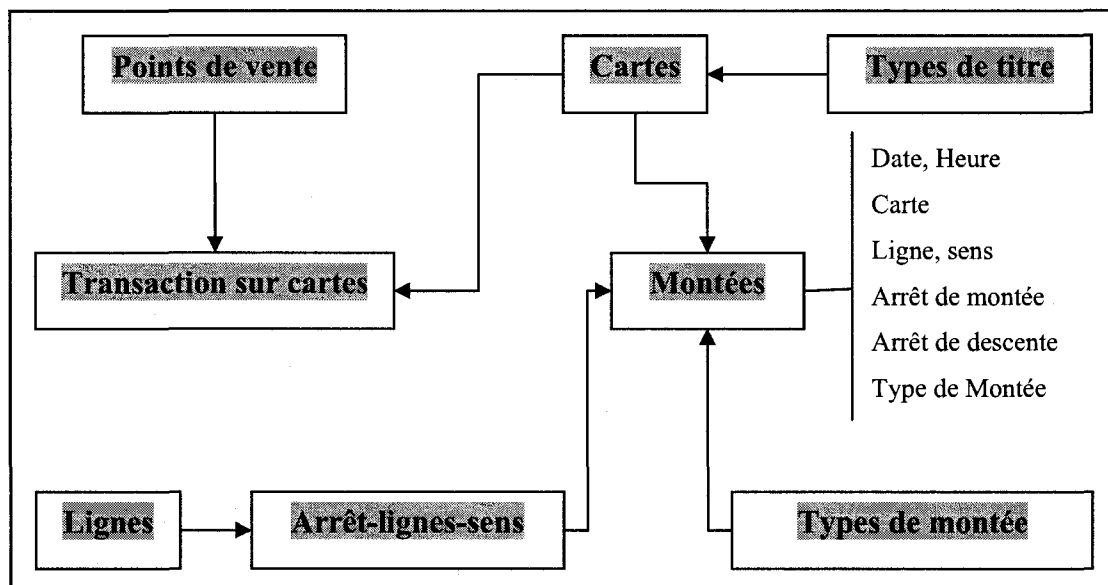


Figure 2.5. Modèle relationnel simplifié

2.3.1. Base de données désagrégée

La base de données désagrégée contient toutes les validations de carte à puce lors d'une montée sur le réseau de transport de Gatineau pendant le mois de septembre 2005. A chaque validation, le système enregistre :

- le numéro de la carte de l'utilisateur

- le type d'abonnement (adulte, aîné, étudiant...)
- le type de voyage (correspondance ou non)
- la date de validation
- l'heure de validation
- l'arrêt, la ligne et le sens de validation
- le numéro du bus et du chauffeur
- l'heure du début de voyage du bus sur cette ligne

On dispose aussi d'une base de données contenant la description des ligne-direction du réseau, avec l'ordre des arrêts et les distances parcourues entre les arrêts

Malheureusement, on n'a aucune information directe sur les destinations des usagers. Cependant, avec toutes ces données, il est possible d'estimer l'arrêt de descente de la plupart des voyages des usagers grâce à un algorithme qui retrace les chaînes de déplacements d'un usager donné (Trépanier et Chapleau, 2006). L'idée de base de l'algorithme est que l'on sait où un usager remonte dans le bus, on peut donc supposer que la destination du trajet précédent était à un arrêt proche de l'arrêt où il remonte. Ces destinations dérivées sont ajoutées à la base de données désagrégée.

Cette base de données contient environ 800 000 transactions pour environ 23 000 cartes différentes pour le mois de septembre 2005. Le réseau est composé d'environ 1 500 arrêts différents et de 107 lignes.

2.3.2. Base de données agrégée

La base de données agrégée détaille pour chaque jour de la période, allant du mois de septembre 2005 au mois de décembre 2005, et pour chaque arrêt-ligne-direction, le nombre de montants par période d'une heure de la journée ainsi que le nombre total des montants pour chaque jour. Le Tableau 2.2 présente la forme de la base de données.

Ainsi, on lit pour l'arrêt 1, sur la ligne 33, dans le sens 0, le 10 septembre 2005, 5 personnes sont montées entre 21h et 22h, 12 personnes sont montées sur toute la journée.

Tableau 2.2. Base de données agrégée

Arrêt	Ligne	Sens	Jour	4h	5h	6h	21 h	22 h	23 h	Total
1	33	0	10-sept-05	0	0	0	5	0	0	12
1	37	0	07-nov-05	0	0	0	0	0	0	1
1	64	0	18-oct-05	0	0	0	0	0	0	1
1	65	1	29-sept-05	0	0	0	0	0	0	40
1	66	0	01-dec-05	0	0	0	0	0	0	1

cette base de données contient près de 300 000 enregistrements. Le réseau de la STO est composé d'environ 6 400 arrêt-ligne-direction différents. L'arrêt où passe le plus de lignes est l'arrêt «5010» où passent 60 ligne-direction différentes. Entre les mois de septembre et décembre 2005, on a enregistré plus de 3 millions de montées sur le réseau de la STO.

2.4. Montage du SIG du réseau de la STO

Pour mener à bien cette étude sur la caractérisation des lieux d'activités, on a besoin d'organiser et de représenter les données par rapport au territoire desservi par la STO. Pour ce faire, nous allons créer un SIG du réseau de la STO. En plus des bases de données décrites précédemment, on dispose d'une base de données contenant la description des arrêts, avec leurs coordonnées et leurs noms, qui va être la base de ce SIG. Le SIG va être complété au fur et à mesure de l'étude avec les différentes caractéristiques étudiées.

2.4.1. Organisation du SIG

Ce SIG est composé de cinq couches :

- les arrêts de bus de la STO (voir 2.4.2)

- les arrêt-ligne-direction (voir 2.4.3)
- les informations agglomérées selon un découpage du territoire (voir 2.4.4)
- les générateurs de déplacements (voir 2.4.5)
- un fond de carte des rues de l'Outaouais

Pour visualiser les cartes qui découlent de ce SIG, on a utilisé le logiciel fGIS (ref).

2.4.2. La couche des arrêts de bus du réseau

Entre septembre 2005 et décembre 2005, l'utilisation du réseau d'autobus de la STO par les utilisateurs de cartes à puce a engendré 1698 arrêts, 111 lignes. La Figure 2.6 montre la répartition des arrêts de la STO sur le territoire de Gatineau. Les arrêts sont très dispersés et leur densité est très variable (très forte dans le centre de Gatineau et moindre dans les lieux résidentiels), ceci rend difficile une quelconque représentation numérique sur ces arrêts.

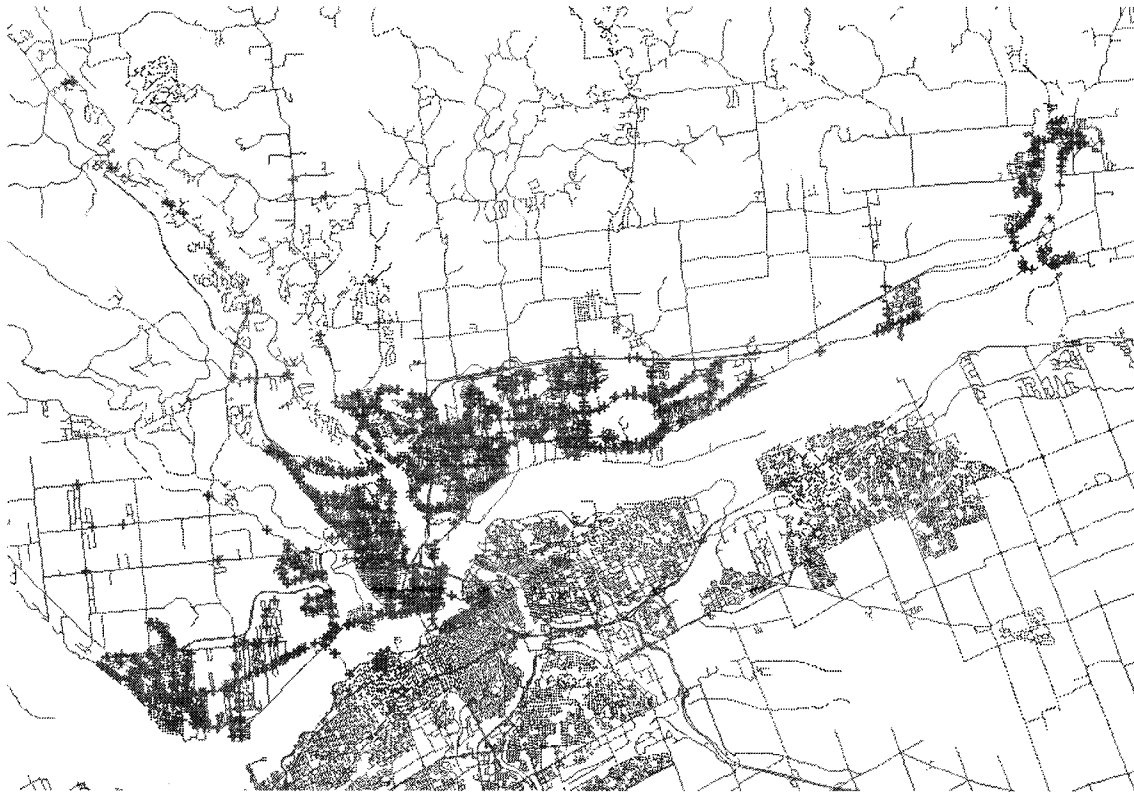


Figure 2.6. Répartition géographique des arrêts de la STO

Cette couche contient pour chaque arrêt :

- ses coordonnées géographiques
- son nom
- la moyenne du nombre de montants par heure de la journée
- le secteur de la grille auquel appartient l'arrêt (voir 2.4.4)
- les informations de proximité avec les générateurs de déplacements (voir 3.4.1)

2.4.3. La couche des arrêt-ligne-direction

Le réseau de transport de la STO est composé de 6536 arrêt-ligne-direction. Cette couche est difficile à représenter car beaucoup d'arrêt-ligne-sens sont superposés, par exemple ceux qui sont issus physiquement du même arrêt.

Cette couche contient pour chaque arrêt-ligne-direction :

- ses coordonnées géographiques
- son nom
- la moyenne du nombre de montants par heure de la journée
- le secteur de la grille auquel appartient l'arrêt (voir 2.4.4)
- les informations issues des analyses de data mining (voir 3.2.1.33.2.1.3)

2.4.4. La grille de regroupement des données

Afin d'améliorer la lisibilité des données, nous avons décidé de créer une grille géographique et de regrouper les arrêts qui sont dans un même secteur de la grille. Ainsi, ces agrégats nous permettent de représenter plus facilement les données. On doit donc choisir une façon de faire ce découpage. Pour notre étude, le territoire de la STO a été arbitrairement découpé en 30 sections dans chaque sens (voir Figure 2.7). Les zones de la grille sont donc rectangulaires, elles mesurent environ 1300m (dans le sens des

« abscisses », direction est-ouest) sur environ 800m (dans le sens des « ordonnées », direction nord-sud).

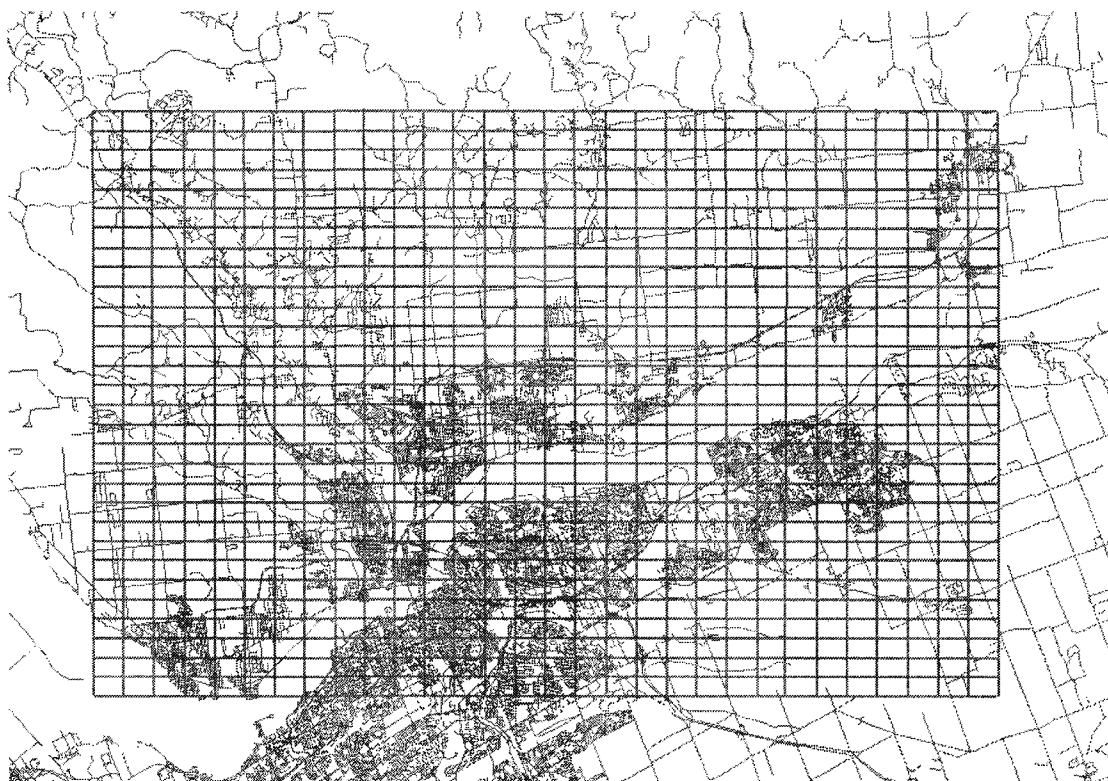


Figure 2.7. Grille de regroupement des données

On peut par exemple représenter les arrêts situés dans chacune des cases de la grille (voir Figure 2.8).

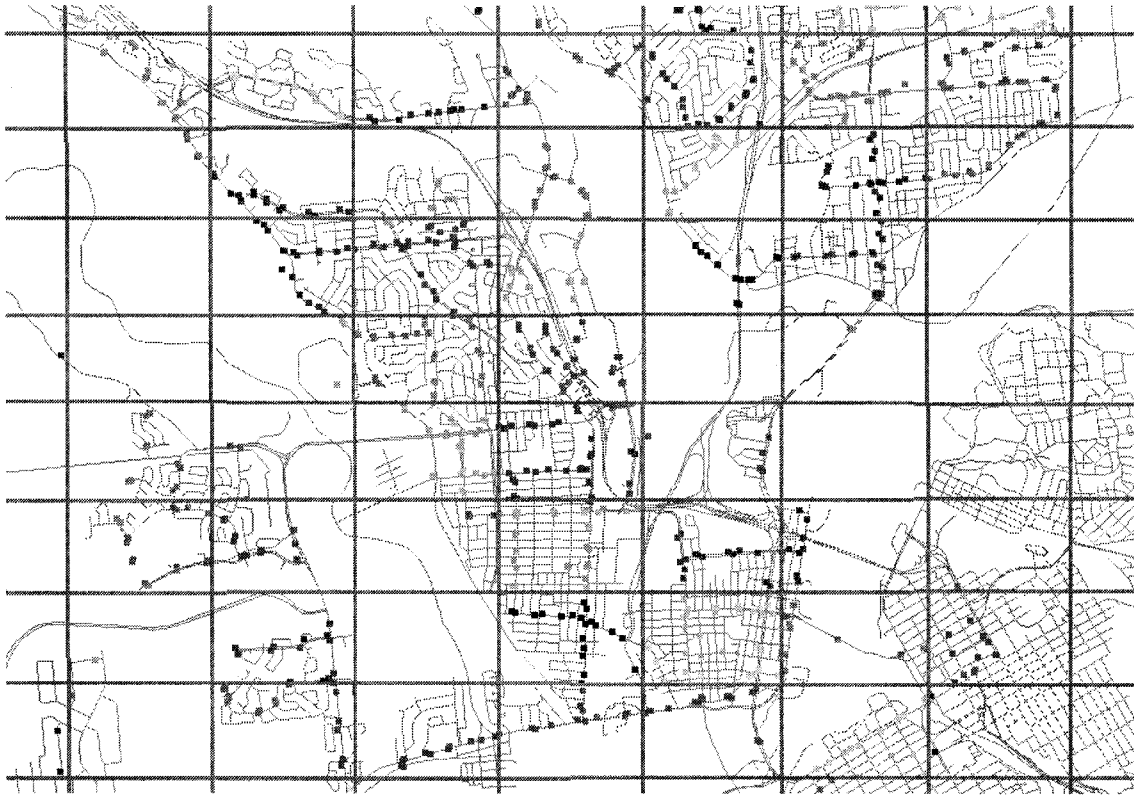


Figure 2.8. Arrêts par case de la grille

Cette couche contient pour chaque case de la grille:

- le nombre moyen de montants par heure de la journée
- le nombre de descendants par heure de la journée
- les informations issues des analyses de data mining (voir 3.1.3)

2.4.5. Les générateurs de déplacements

Grâce aux sites Internet institutionnels de la STO, de la ville de Gatineau et de la commission scolaire de l'Outaouais, on identifie les principaux lieux d'activités de l'Outaouais en recherchant au moins leurs noms et adresses. Ces lieux d'activités sont les générateurs de déplacements les plus importants du réseau de la STO.

Les lieux d'activités principaux de l'Outaouais sont :

- les écoles secondaires (11 au total sur le territoire étudié), les cégeps (4) et les universités (3)

- les centres commerciaux (4); au total quatre grands centres commerciaux dans l'Outaouais où les gens vont magasiner, on ne compte pas ici les centres commerciaux dans le centre ville pour éviter de fausses analyses
- les hôpitaux (4)
- le centre ville d'Ottawa et de Gatineau : cette zone est le plus grand générateur de déplacements, le travail étant la principale raison de déplacements. Le centre ville est composé de tours de bureaux mais aussi de magasins et de lieux de culture (cinémas, théâtres...)
- les stationnements incitatifs (12) : la STO met à la disposition de sa clientèle douze stationnements incitatifs (parc-o-bus) gratuits répartis sur tout le territoire
- les domiciles des usagers : ils sont disséminés sur tout le territoire, pas de grande concentration. Il est difficile de mesurer leur influence, on ne les étudie pas dans ce mémoire

Dans le réseau de transport en commun de la STO, il existe des arrêt-ligne-direction sur lesquels les usagers effectuent beaucoup de changements de bus. Ils arrivent à ces *nœuds* du réseau par une ligne donnée, y font une correspondance et repartent par une autre ligne. La base de données désagrégée nous permet de savoir si un voyage effectué par un usager est une correspondance ou non.

On appelle *Nœud du Réseau* un arrêt-ligne-direction sur lequel plus de 300 usagers font une correspondance pour tout le mois de septembre. On compte donc pour chaque arrêt-ligne-direction le nombre de correspondances pour le mois de septembre. Le seuil de 300 usagers est fixé arbitrairement. Sur les 6536 arrêt-ligne-direction du réseau, on obtient 81 nœuds.

La carte suivante (Figure 2.9) présente les arrêt-ligne-direction qui sont des nœuds du réseau. On ne voit pas les 81 arrêt-ligne-direction car beaucoup sont situés au même arrêt.



Figure 2.9. Les nœuds du réseau

On ajoute à cette liste les nœuds du réseau qui sont des arrêts très actifs et dont l'activité n'est due qu'à l'exploitation du réseau. Ces nœuds ne sont pas forcément des lieux d'activités mais ils influencent beaucoup l'utilisation du réseau car ce sont des générateurs pour beaucoup de sections de déplacements.

Ensuite, on localise géographiquement (par coordonnées x,y) tous ces lieux d'activités et on construit une base de données géographique. Pour ce faire, on utilise les outils Internet Googlemaps.ca et Wikimapia.org. La carte suivante (Figure 2.10) présente la localisation des générateurs de déplacements identifiés (sauf les nœuds qui rendaient illisibles la carte).

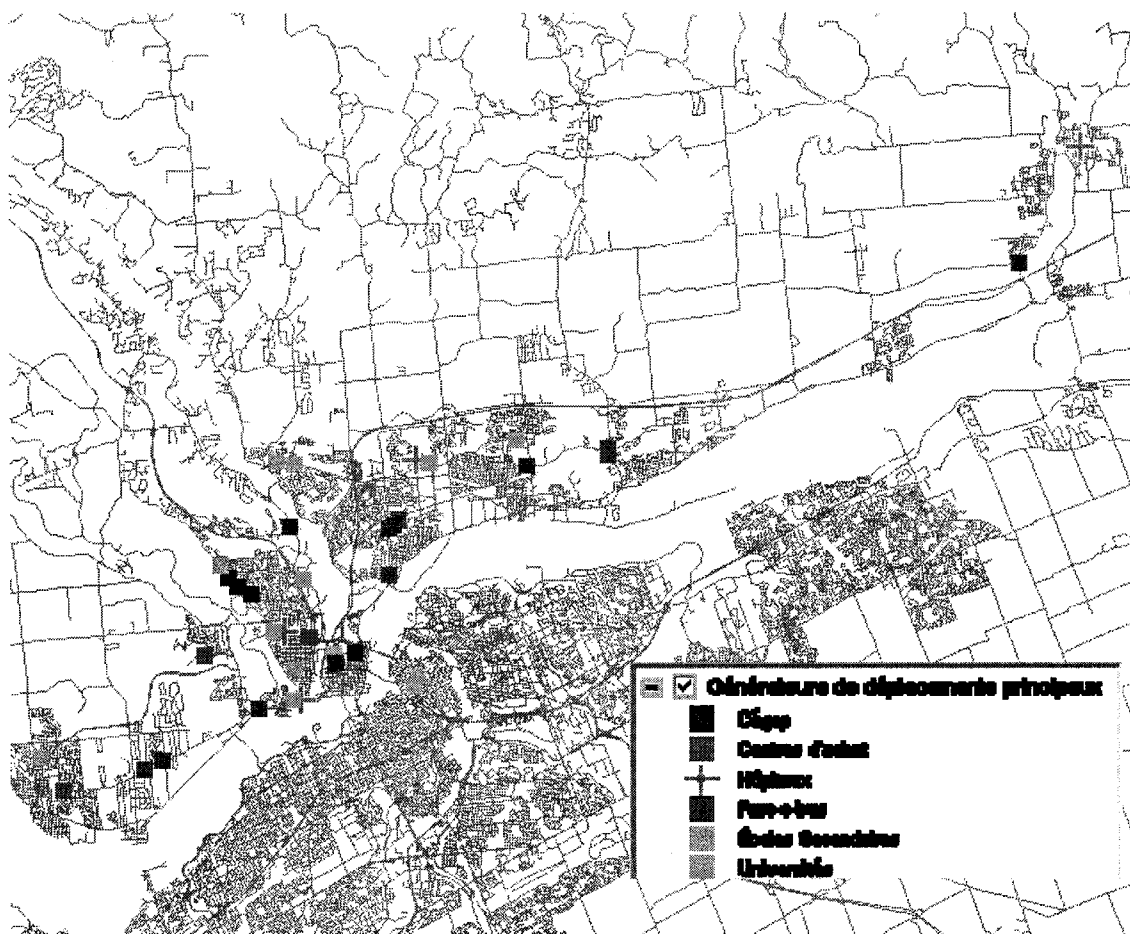


Figure 2.10. Les principaux générateurs de déplacements

2.5. Tendances principales

2.5.1. Le profil moyen des montants

Les données de carte à puce permettent de confirmer les variations bien connues du profil des montants pendant une journée. Pour chaque jour de la semaine, du lundi au dimanche, on calcule le nombre de montants par heure (la base de données agrégée entre septembre et décembre 2005 contient quasiment le même nombre de chaque jour, 17 dimanches, lundis, mardis et mercredis et 18 jeudis, vendredis et samedis).

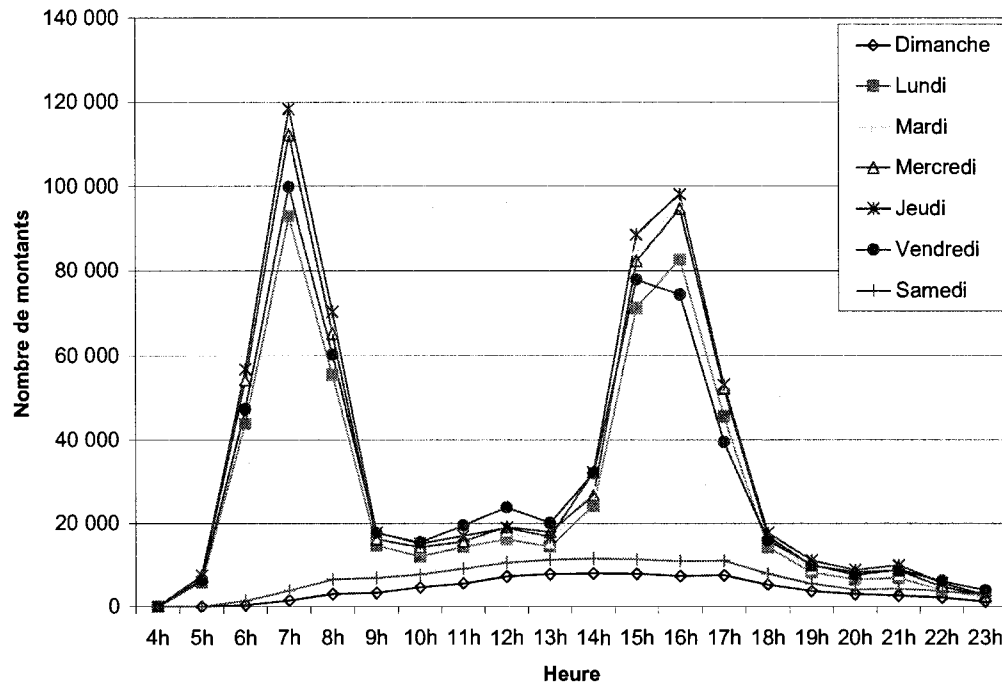


Figure 2.11. Profil des montants par heure et par jour de la semaine (somme sur toute la période d'observation)

Lorsqu'on regarde la Figure 2.11, on peut tout d'abord observer une grande différence entre la semaine (du lundi au vendredi) et la fin de semaine. Le réseau de transport est beaucoup moins utilisé le week-end car les usagers utilisent le transport en commun surtout pour aller travailler.

Pendant la semaine, on voit nettement apparaître sur les courbes les deux périodes de pointe (le matin entre 6h et 9h et la fin d'après-midi entre 15h et 18h), la période de pointe du matin étant plus courte que celle de l'après-midi. Cela est dû au fait que la journée typique de travail d'adulte en Outaouais commence vers 9h et finit vers 16h30. On constate aussi que mardi, mercredi et jeudi sont les jours où il y a le plus d'utilisateurs sur le réseau. Puis viennent les vendredis et les lundis. Le vendredi est singulier, les usagers ont tendance à utiliser les transports en commun plus tôt pendant l'heure de pointe de l'après-midi, probablement parce que les gens y terminent leur journée de travail plus tôt que le reste de la semaine.

Pendant la fin de semaine, il y a plus d'usagers les samedis que les dimanches et les usagers utilisent le réseau plus tard le matin que pendant la semaine. Le week-end, la charge des montants est plus régulière tout au long de la journée.

2.5.2. Nombre d'arrêt-ligne-direction utilisés par jour

La figure suivante présente le nombre d'arrêt-ligne-direction utilisés par jour pendant toute la période étudiée.

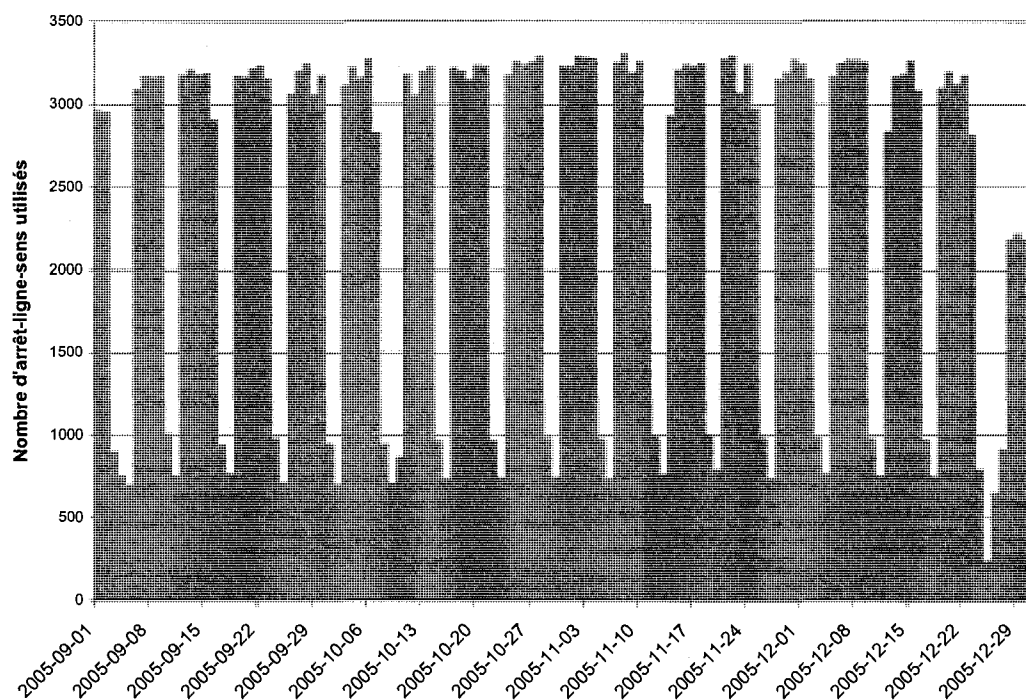


Figure 2.12. Nombre d'arrêt-ligne-direction utilisés par jour

On voit clairement apparaître les variations hebdomadaires (Figure 2.12). Le week-end, le réseau est beaucoup moins utilisé que pendant la semaine. On voit aussi apparaître les jours fériés (Fête du travail, Action de Grâce, fêtes de fin d'année). On note aussi la similitude du patron d'une semaine sur l'autre.

2.5.3. Origine-destination

La Figure 2.13 et la Figure 2.14 présentent le nombre de montants dans chaque zone et la destination de ces montants entre 8h et 9h (Figure 2.13) et entre 16h et 17h (Figure 2.14) sur tout le mois de septembre 2005.

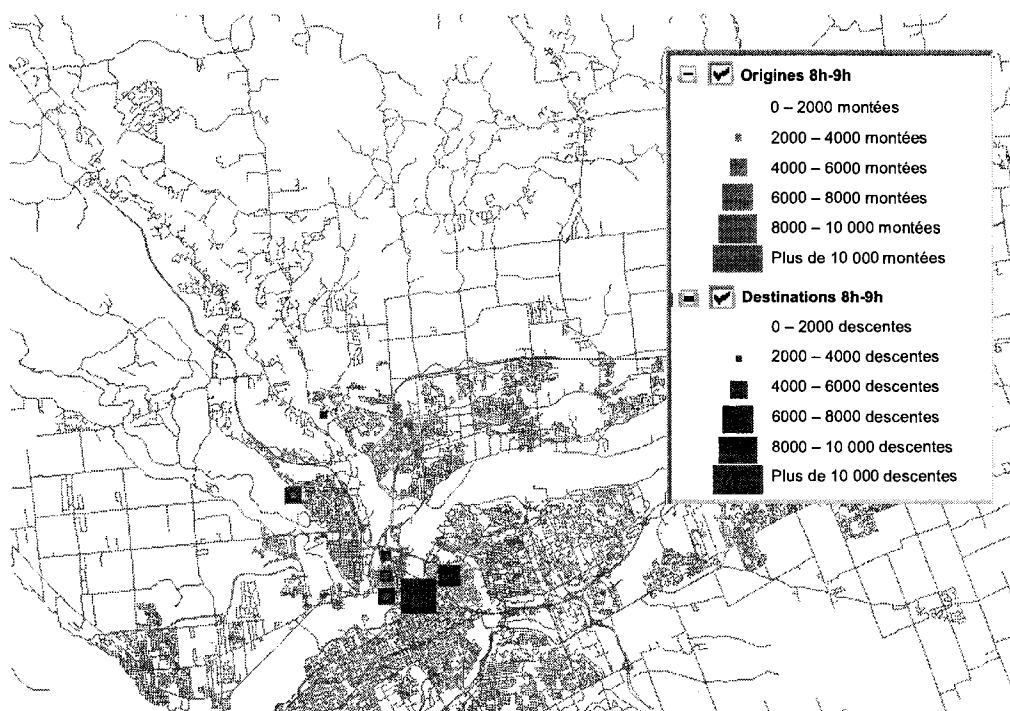


Figure 2.13. Origines (rouge) et destinations (bleu) des montants entre 8h et 9h dans chaque case de la grille pendant le mois de septembre 2005

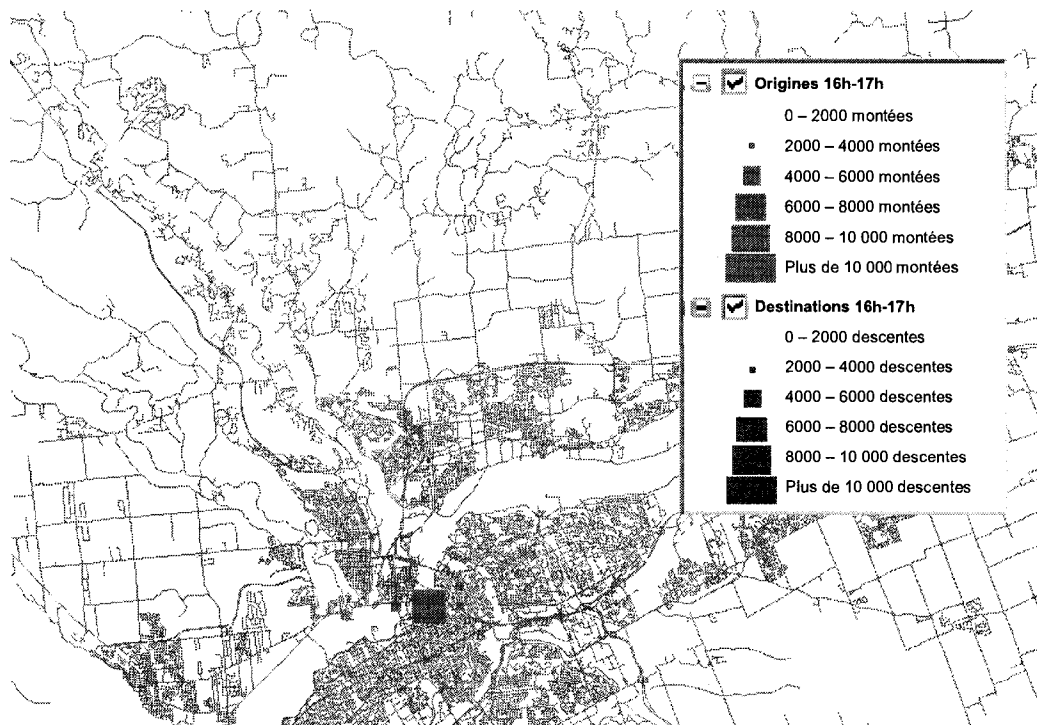


Figure 2.14. Origines (rouge) et destinations (bleu) des montants entre 16h et 17h dans chaque case de la grille pendant le mois de septembre 2005

Ces cartes (Figure 2.13 et Figure 2.14) nous permettent de voir la prépondérance des centres-villes de Gatineau et d'Ottawa qui attirent beaucoup d'utilisateurs à l'heure de pointe du matin et d'où partent la majorité des usagers pendant la période de pointe de l'après-midi. Cela correspond aux voyageurs qui vont travailler et qui rentrent chez eux en fin de journée.

2.5.4. Distance parcourue lors d'une section de déplacement

Grâce aux destinations dérivées de chaque section de déplacement, on peut estimer la distance parcourue par un utilisateur à chaque fois qu'il monte dans le bus. On peut ensuite faire la moyenne des distances parcourues pour chaque secteur de la grille et on obtient la carte suivante.

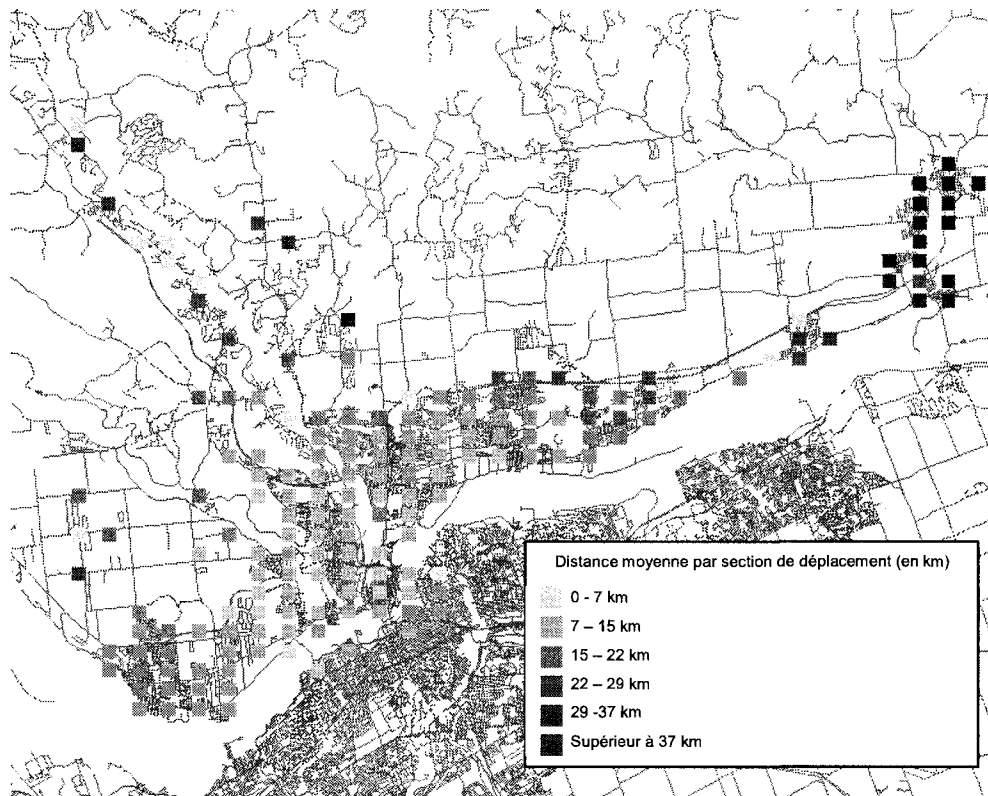


Figure 2.15. Distance moyenne (en mètre) parcourue au cours d'une section de déplacement au départ des différentes zones

On peut remarquer (Figure 2.15) que les zones les plus sombres (celles où les usagers font le plus de kilomètres) sont celles qui sont les plus éloignées du centre ville de Gatineau. La distance moyenne dans ces secteurs est quasiment égale à la distance du secteur au centre ville. Cela veut dire qu'un utilisateur moyen dans ces zones utilise le bus surtout pour aller au centre ville.

2.5.5. Les classes d'utilisateurs

On a vu en 2.1.4 qu'il existe plusieurs sortes d'abonnement au système de transport de la STO et que cela permet d'identifier les types d'utilisateurs : adultes, aînés, étudiants ou employés de la STO. Dans cette section, nous allons étudier les comportements des usagers selon leur catégorie.

Le Tableau 2.3 présente le nombre et le pourcentage de cartes par classe ainsi que le nombre et le pourcentage de montées par classe.

Tableau 2.3. Répartition des usagers dans les classes

Classe	Nombre de cartes	Pourcentage de cartes	Nombre de montées	Pourcentage de montées	Nombre moyen de montées par usagers
Adulte	13 029	57,26%	477 423	58,41%	36,6
Étudiant	8 708	38,27%	308 613	37,76%	35,4
Ainé	827	3,63%	30 141	3,69%	36,4
STO	189	0,83%	1 122	0,14%	5,9
Total	22 753	100,00%	817 299	100,00%	35,9

On peut voir que pour les trois types de cartes principaux, les usagers font quasiment le même nombre moyen de montées pendant le mois de septembre 2005. Le graphique suivant (Figure 2.16) présente pour chaque classe d'usagers le profil des montants au cours de la journée. On a vu en qu'il y avait beaucoup de différences entre les jours de la semaine et le week-end. Pour être en mesure de comparer, on prend en compte les données de montées sur tous les jeudis du mois de septembre uniquement.

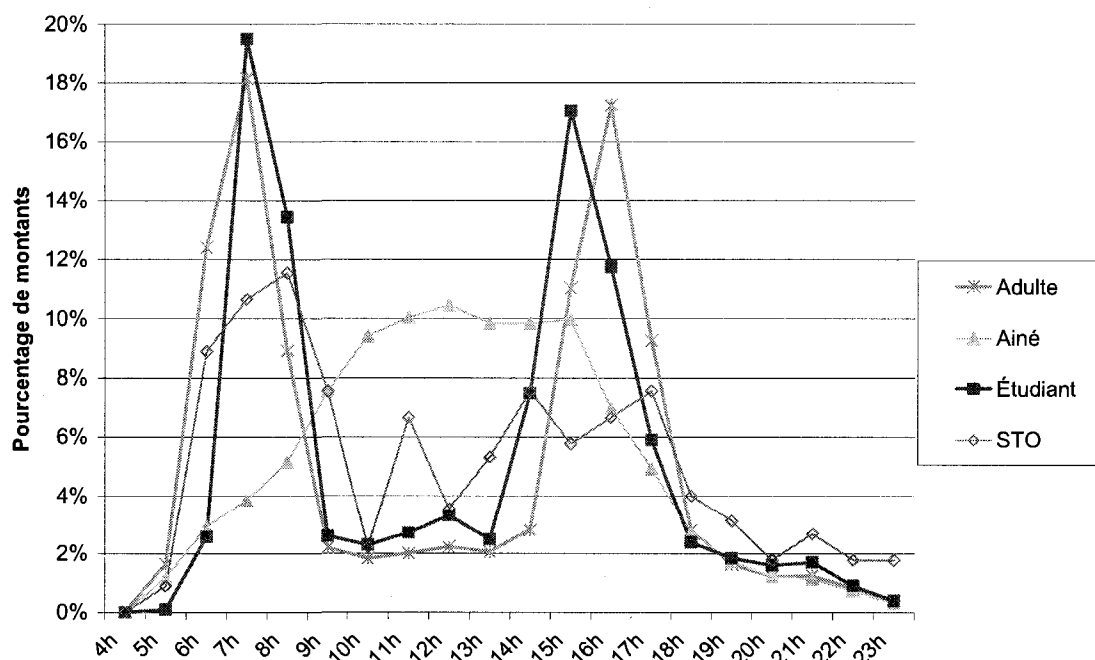


Figure 2.16. Profil des montants par classe pour les jeudis de septembre 2005

La figure précédente (Figure 2.16) se lit ainsi : environ 20% des montants adultes montent entre 7h et 8h du matin. Cette figure permet de constater les différences de comportements entre les classes d'usagers. Effectivement, pendant la semaine, les adultes ont tendances à utiliser les transports en commun plus tôt le matin et plus tard le soir que les étudiants et les aînés voyages surtout au milieu de la journée entre 10h et 16h.

La figure suivante (Figure 2.17) présente le pourcentage de montées pour chaque type d'usagers par heure de la journée pour tous les jeudis du mois de septembre 2005. La ligne rouge montre la proportion moyenne d'usagers adultes.

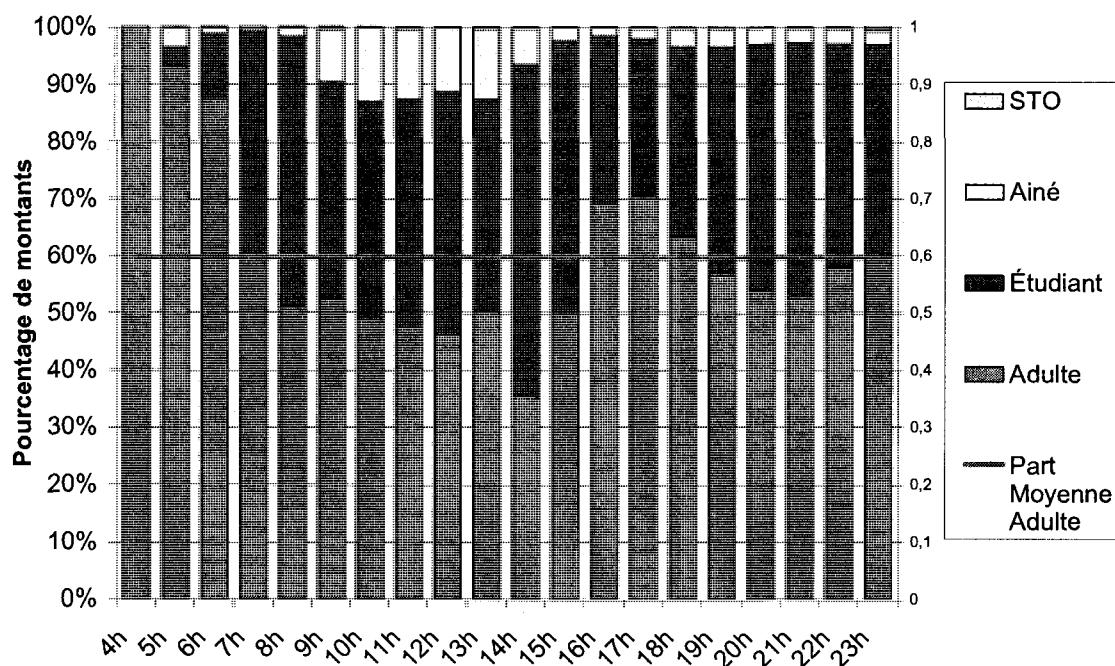


Figure 2.17. Proportion des classes d'usagers par heure pour tous les jeudis de septembre 2005

La figure précédente permet de constater que les adultes sont majoritaires surtout pendant les heures de pointes (entre 6h et 8h du matin et entre 16h et 19h le soir). Entre 8h et 9h et entre 14h et 16h, ce sont les étudiants qui utilisent principalement le réseau.

Suite à ces éléments descriptifs, nous allons passer à une analyse approfondie des comportements liés aux arrêts du réseau.

CHAPITRE 3. EXPÉRIMENTATION ET RÉSULTATS

3.1. Segmentation en groupes d'arrêt-ligne-direction homogènes

Dans cette partie, nous allons étudier l'objet arrêt-ligne-direction. Pour ce faire, nous voulons créer des groupes d'arrêt-ligne-direction qui ont des comportements semblables en termes d'utilisation. Dans un premier temps, nous allons préparer les données issues de la base de données agrégée. Dans un second temps, nous allons déterminer un nombre adéquat de groupes d'arrêt-ligne-direction homogènes grâce à des techniques de data mining.

3.1.1. Préparation des données

Tout d'abord, les journées sont segmentées en intervalles d'une durée d'une heure en commençant à 4h du matin. Ensuite, on a agrégé pour chaque arrêt-ligne-direction et pour chaque jour des mois de septembre 2005 à décembre 2005, le nombre de montants dans chaque intervalle de la journée (Tableau 2.2). Puis, on normalise chaque ligne en divisant par le total de chaque ligne, pour avoir la distribution des montants par heure de la journée sur chaque arrêt-ligne-direction pour chaque jour de la période considérée.

Tableau 3.1. Proportion de montants par heure, par arrêt-ligne-direction et par jour

[illegible]

Finalement, on obtient une base de données avec le taux d'utilisation de chaque arrêt-ligne-direction pendant chaque période d'une heure de chaque journée (cf. Tableau 3.1).

3.1.2. Fouille de données et sélection d'un nombre adéquat de clusters

Maintenant, nous voulons créer des groupes de comportements homogènes dans notre échantillon d'étude. L'objectif est de révéler des groupes naturels d'arrêt-ligne-direction-jour. Pour ce faire, on utilise le logiciel Tanagra. On met comme paramètres d'entrée pour le traitement les proportions de montants par période de la journée (Moy_4h à Moy_23h). Étant donné le grand nombre de données dans la base de départ, le temps de calcul pour un algorithme de segmentation par agglomération hiérarchique sur toute la base de données est prohibitif, on fait donc tourner cet algorithme sur des sous-groupes calculés avec l'algorithme k-means. Le nombre « optimal » de clusters donné par l'algorithme par agglomération hiérarchique est de 21. On obtient le dendrogramme de la Figure 3.1, où on voit la marche pour 21 groupes.

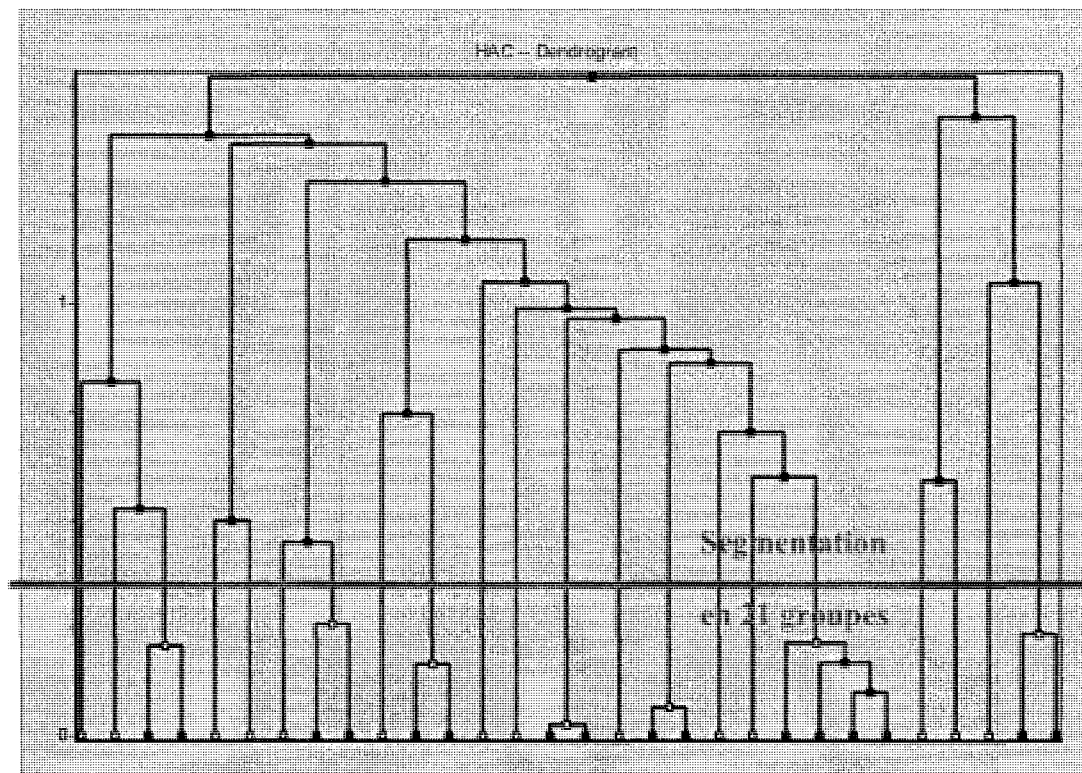


Figure 3.1. Dendrogramme fait avec 30 sous-groupes.

Puis, on fait tourner un algorithme k-means sur toute la base de données pour faire 21 groupes. On obtient la répartition suivante des arrêt-ligne-direction dans les différents groupes (cf. Tableau 3.2).

Tableau 3.2. Groupe d'arrêt-ligne-direction (en ordre décroissant du nombre de montants)

Cluster	Nombre d'arrêt-ligne-direction-jour	Nombre de montants	Pourcentage d'arrêt-ligne-direction-jour	Pourcentage de montants
c_kmeans_9	47 965	969 785	16,42%	29,83%
c_kmeans_12	42 940	594 313	14,70%	18,28%
c_kmeans_16	21 733	490 242	7,44%	15,08%
c_kmeans_15	16 406	325 067	5,62%	10,00%
c_kmeans_14	34 967	176 485	11,97%	5,43%
c_kmeans_7	18 179	131 908	6,22%	4,06%
c_kmeans_17	23 308	124 548	7,98%	3,83%
c_kmeans_6	10 058	108 763	3,44%	3,35%
c_kmeans_18	12 810	89 050	4,38%	2,74%
c_kmeans_20	10 909	48 751	3,73%	1,50%
c_kmeans_21	4 984	41 895	1,71%	1,29%
c_kmeans_8	7 063	25 530	2,42%	0,79%
c_kmeans_3	8 147	23 667	2,79%	0,73%
c_kmeans_4	6 467	19 384	2,21%	0,60%
c_kmeans_19	3 953	17 080	1,35%	0,53%
c_kmeans_1	5 890	16 147	2,02%	0,50%
c_kmeans_13	5 023	13 302	1,72%	0,41%
c_kmeans_10	3 672	12 857	1,26%	0,40%
c_kmeans_2	4 209	11 775	1,44%	0,36%
c_kmeans_5	2 019	6 378	0,69%	0,20%
c_kmeans_11	1 439	4 302	0,49%	0,13%
Total	292 141	3 251 229	100,00%	100,00%

En comparant les pourcentages de nombre d'arrêt-ligne-direction-jour et de montants, on s'aperçoit que ces deux notions ne sont pas proportionnelles. Ainsi, le groupe 9 regroupe

16% des arrêts et 30% des montants, ce groupe est composé d'arrêts sur lesquels il y a beaucoup de montants. A l'opposé, le groupe 14 regroupe environ 12% des arrêts mais seulement 5% des montants.

Les quatre principaux groupes (en termes de nombre de montants), les groupes 9, 12, 16 et 15, récoltent au total près de 75% des montants.

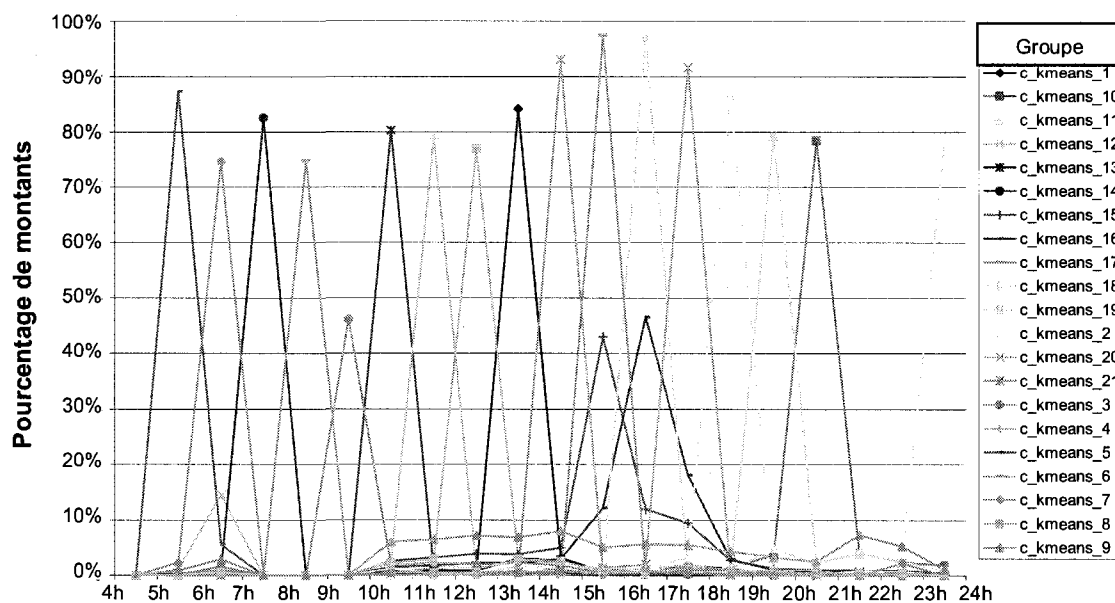


Figure 3.2. Profil de la distribution des montants en fonction de l'heure de la journée pour 21 clusters

Dans la Figure 3.2, on lit le pourcentage (sur toute la journée) d'utilisateurs qui montent pendant une période donnée. Ainsi, environ 87% des montées sur les arrêt-ligne-direction du groupe 5 se font entre 5h et 6h.

Malheureusement, avec 21 groupes, les données sont difficilement exploitables. En effet, 12 groupes rassemblent chacun moins de 2% des montants. Comme ces groupes ne sont pas significatifs, il faudrait les regrouper avec d'autres groupes pour avoir une meilleure lisibilité. Dès lors, puisque les 21 groupes optimaux d'un point de vue mathématique ne nous permettent pas d'avoir une bonne compréhension de l'utilisation

des arrêt-ligne-direction, on diminue le nombre de groupes. Pour ce faire, on peut remarquer que 6 des 12 groupes les plus petits ont des pics d'utilisation consécutifs au milieu de la journée entre 9h du matin et 15h. Pour éviter d'avoir peu d'arrêt-ligne-direction dans beaucoup de groupes, on va changer la base des temps en agrégeant tous les montants d'entre 9h et 15h dans un même paquet, appelé « heure creuse du milieu de journée ». Ainsi, tous ces arrêt-ligne-direction seront agglomérés dans un groupe qui aura un pic au milieu de la journée (voir Tableau 3.3). Finalement, l'heure importera peu pour le milieu de la journée.

Tableau 3.3. Nouveau découpage temporel d'une journée avec un regroupement des heures du milieu de la journée

	Heures de la journée																							
Initialement	4h	5h	6h	7h	8h	9h	10h	11h	12h	13h	14h	15h	16h	17h	18h	19h	20h	21h	22h	23h				
Regroupement du milieu de la journée	4h	5h	6h	7h	8h	Heures creuses du milieu de la journée							15h	16h	17h	18h	19h	20h	21h	22h	23h			

Ensuite, on recommence les algorithmes de data mining sur la nouvelle base de données où le milieu de la journée est agrégé.

Avec l'algorithme de segmentation par agglomération hiérarchique fait sur des sous-groupes, on obtient le dendrogramme suivant (Figure 3.3) :

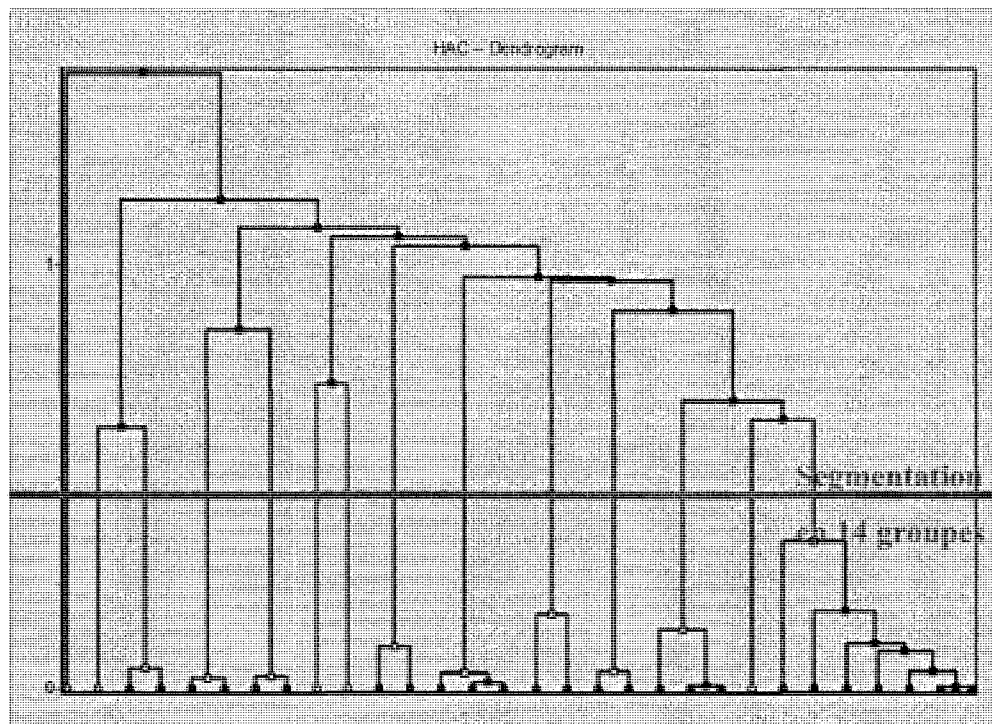


Figure 3.3. Dendrogramme fait avec 30 sous-groupes

L'algorithme recommande de faire 14 clusters, c'est moins que les 21 clusters recommandés avec le découpage complet par heure de la journée.

Ensuite, on applique l'algorithme k-means sur la base de données pour faire 14 groupes. On obtient la répartition suivante dans les différents clusters (Tableau 3.4) :

Tableau 3.4. Groupe d'arrêt-ligne-direction (en ordre décroissant du nombre de montants)

Cluster	Nombre d'arrêt-ligne-direction-jour	Nombre de Montants	Pourcentage d'arrêt-ligne-direction-jour	Pourcentage de montants
c_kmeans_3	57 950	1 135 689	19,84%	34,93%
c_kmeans_12	24 172	597 246	8,27%	18,37%
c_kmeans_10	39 549	517 007	13,54%	15,90%
c_kmeans_4	41 260	204 898	14,12%	6,30%
c_kmeans_14	34 794	173 371	11,91%	5,33%
c_kmeans_1	12 853	141 500	4,40%	4,35%
c_kmeans_5	18 208	133 841	6,23%	4,12%
c_kmeans_13	23 004	122 892	7,87%	3,78%
c_kmeans_2	14 154	108 101	4,84%	3,32%
c_kmeans_7	10 906	47 503	3,73%	1,46%
c_kmeans_9	5 439	33 015	1,86%	1,02%
c_kmeans_6	3 462	16 971	1,19%	0,52%
c_kmeans_11	4 219	12 136	1,44%	0,37%
c_kmeans_8	2 171	7 059	0,74%	0,22%
Total	292 141	3 251 229	100,00%	100,00%

La Figure 3.4 présente la proportion de montants par période de la journée.

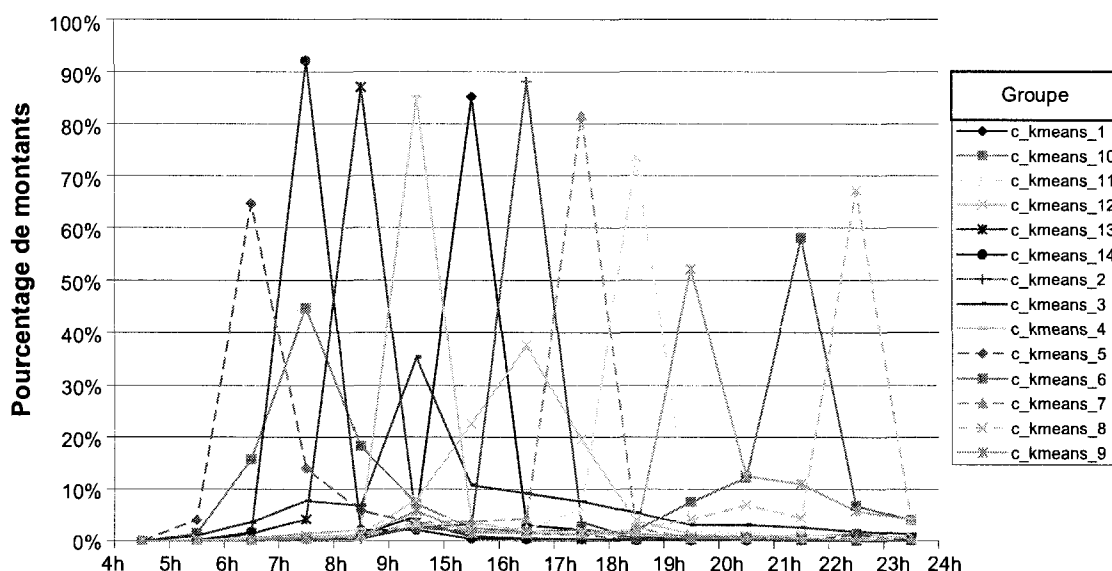


Figure 3.4. Profil de la distribution des montants en fonction de la période de la journée pour 14 clusters

Même si on a réussi à diminuer le nombre de clusters en passant de 21 à 14, on a toujours trop de clusters pour pouvoir faire une analyse claire et il reste encore cinq clusters qui représentent moins de 2% des montants. Ces petits clusters ne nous donnent pas des tendances suffisamment significatives pour pouvoir tirer des conclusions. Mais, chose étonnante, cette fois encore les pics d'activités de ces petits clusters sont consécutifs. Le groupe 7 a son pic de proportion des montants entre 17h et 18h, le groupe 11 entre 18h et 19h, le groupe 9 entre 19h et 20h, le groupe 6 entre 21h et 22h et le groupe 8 entre 23h et 24h, on peut noter qu'aucun groupe d'arrêt-ligne-direction n'a de pic d'activité entre 20h et 21h. Cependant, le groupe 12 possède plus de 20% de ses montants (soit plus de 100 000 montées) entre 17h et 18h. Cette tranche horaire est donc caractéristique d'une partie assez importante des arrêt-ligne-direction, il faut la garder pour avoir un bon niveau de détail. Nous allons donc pouvoir agréger les usagers qui montent sur le réseau entre 18h et 24h dans un groupe appelé «Soir» (voir Tableau 3.5).

Tableau 3.5. Nouveau découpage temporel d'une journée avec un regroupement des heures du milieu de la journée et du soir

	Heures de la journée																						
Initialement	4h	5h	6h	7h	8h	9h	10h	11h	12h	13h	14h	15h	16h	17h	18h	19h	20h	21h	22h	23h			
Regroupement du milieu de la journée						Heures creuses du milieu de la journée																	
	4h	5h	6h	7h	8h							15h	16h	17h	18h	19h	20h	21h	22h	23h			
Regroupement du Soir						Heures creuses du milieu de la journée									Soir								
	4h	5h	6h	7h	8h							15h	16h	17h									

On obtient donc une nouvelle base de données avec la proportion de montants pour chaque arrêt-ligne-direction-jour et pour chaque période de la journée du découpage défini dans le tableau précédent. Pour connaître la meilleure segmentation de la base, on applique un algorithme de segmentation par agglomération hiérarchique.

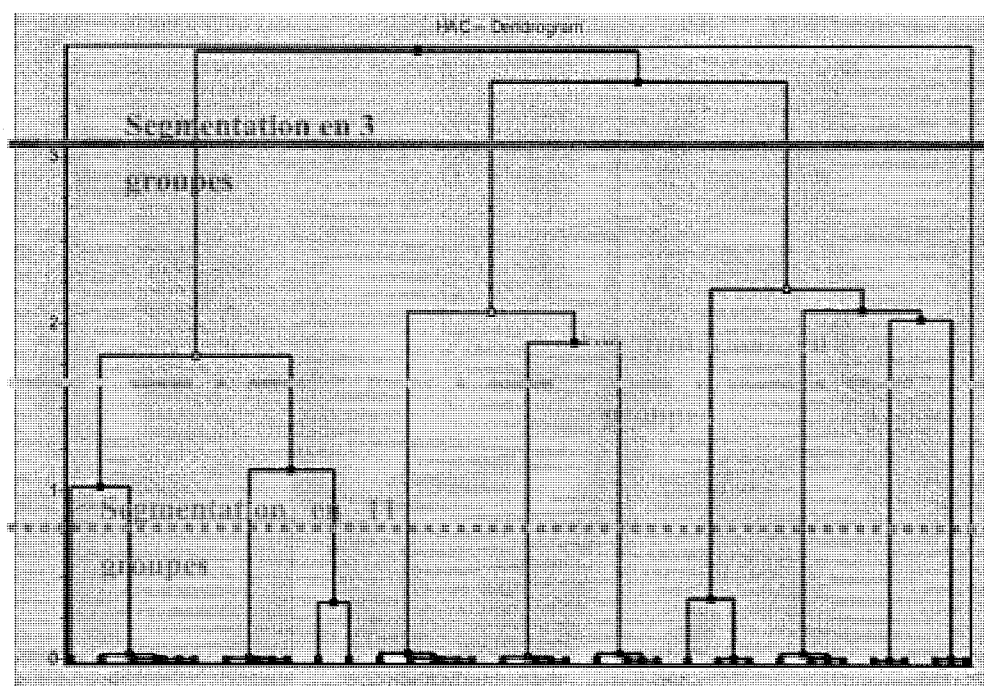


Figure 3.5. Dendrogramme fait avec 30 sous-groupes.

Cet algorithme recommande un nombre optimal de clusters de 3 (voir Figure 3.5). Ensuite, on fait tourner un algorithme de segmentation k-means sur la base de données pour avoir 3 clusters, on obtient la répartition suivante dans les clusters (Tableau 3.6):

Tableau 3.6. Groupe d'arrêt-ligne-direction (en ordre décroissant du nombre de montants)

Cluster	Nombre d'arrêt-ligne-direction-jour	Nombre de Montants	Pourcentage d'arrêt-ligne-direction-jour	Pourcentage de montants
c_kmeans_1	156 062	2 106 643	53,42%	64,80%
c_kmeans_2	70 541	596 863	24,15%	18,36%
c_kmeans_3	65 538	547 723	22,43%	16,85%
Total	292 141	3 251 229	100,00%	100,00%

La figure suivante (Figure 3.6) présente la distribution des montants par période de la journée pour les trois groupes.

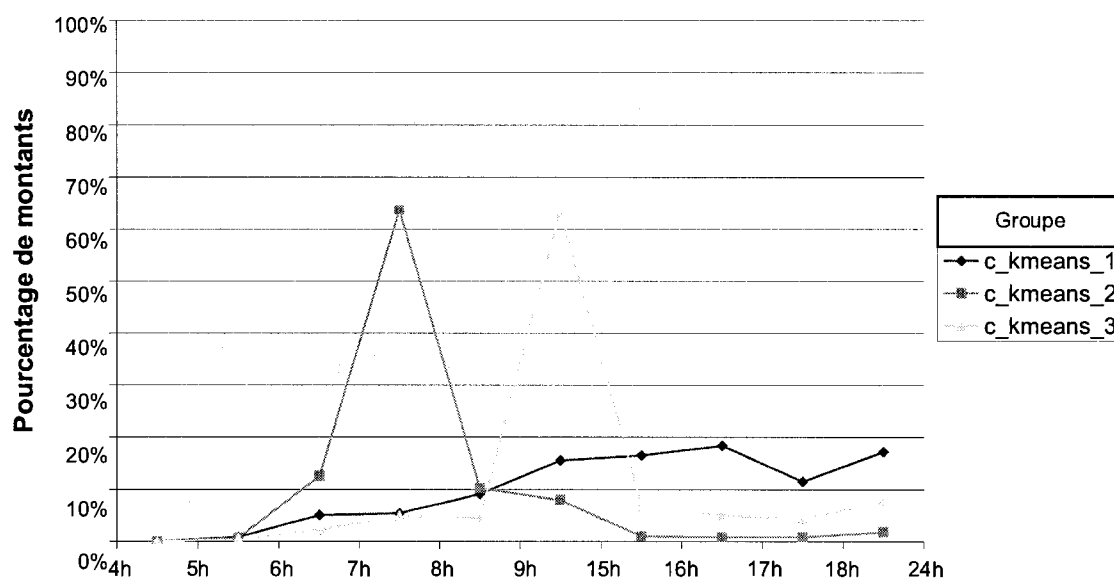


Figure 3.6. Profil de la distribution des montants en fonction de la période de la journée pour 3 clusters

Le groupe le plus important en termes de montants est le groupe 1. Dans ce groupe, les arrêt-ligne-direction sont utilisés régulièrement toute la journée, il n'y a pas de pic d'utilisation. Le groupe 2 est composé d'arrêt-ligne-direction qui sont essentiellement utilisés pendant la période de pointe du matin (plus de 85% des montées sur les arrêt-ligne-direction du groupe se fait entre 6h et 9h). Pour les arrêt-ligne-direction du groupe 3, on observe un pic d'utilisation (plus de 60% des montants) pendant la période du milieu de la journée.

Malheureusement, diviser la population des arrêt-ligne-direction-jour en seulement trois groupes est problématique, on perd beaucoup de définition et la plupart des arrêt-ligne-direction se retrouvent dans un groupe moyen dont l'utilisation se fait assez régulièrement sur toute la journée et dont on ne peut pas dire grand-chose...

On peut donc conclure qu'il faut faire un compromis sur le nombre de groupes. Avec trop de clusters, on est perdu dans les données et beaucoup de groupes ne sont pas significatifs en termes de taille mais avec trop peu de clusters, on perd de la définition et la plupart des arrêt-ligne-direction sont semblables à quelques arrêts moyens.

Si on en revient au dendrogramme (Figure 3.5), on s'aperçoit qu'il y a deux autres marches importantes qui correspondent à des segmentations intéressantes, une pour une segmentation en 9 clusters et une autre pour une segmentation en 11 groupes. On fait donc tourner des algorithmes k-means pour créer 11 et 9 groupes dans la base. Avec 11 groupes, on obtient la répartition suivante dans les groupes (Tableau 3.7).

Tableau 3.7. Groupe d'arrêt-ligne-direction (en ordre décroissant du nombre de montants)

Cluster	Nombre d'arrêt-ligne-direction-jour	Nombre de Montants	Pourcentage d'arrêt-ligne-direction-jour	Pourcentage de montants
c_kmeans_2	49 723	815 436	17,02%	25,08%
c_kmeans_1	35 783	706 062	12,25%	21,72%
c_kmeans_9	40 120	537 069	13,73%	16,52%
c_kmeans_8	26 806	373 551	9,18%	11,49%
c_kmeans_10	34 801	173 471	11,91%	5,34%
c_kmeans_11	15 120	171 761	5,18%	5,28%
c_kmeans_7	34 837	153 994	11,92%	4,74%
c_kmeans_3	18 183	135 011	6,22%	4,15%
c_kmeans_5	23 218	123 490	7,95%	3,80%
c_kmeans_6	11 532	55 103	3,95%	1,69%
c_kmeans_4	2 018	6 281	0,69%	0,19%
Total	292 141	3 251 229	100,00%	100,00%

Ici encore, les deux derniers groupes représentent moins de 2% des montants chacun. Ils ne sont pas vraiment significatifs. Cette segmentation en 11 groupes n'est donc pas vraiment satisfaisante.

Nous allons passer à l'étude de la segmentation en 9 groupes. Le Tableau 3.8 montre la répartition dans chaque groupe pour la segmentation en 9 groupes:

Tableau 3.8. Groupe d'arrêt-ligne-direction (en ordre décroissant du nombre de montants)

Cluster	Nombre d'arrêt-ligne-direction-jour	Nombre de Montants	Pourcentage d'arrêt-ligne-direction-jour	Pourcentage de montants
c_kmeans_4	62 075	1 316 572	21,25%	40,49%
c_kmeans_9	30 862	493 062	10,56%	15,17%
c_kmeans_8	52 634	378 121	18,02%	11,63%
c_kmeans_3	32 111	318 566	10,99%	9,80%
c_kmeans_7	41 300	206 219	14,14%	6,34%
c_kmeans_6	16 392	202 375	5,61%	6,22%
c_kmeans_5	25 141	147 209	8,61%	4,53%
c_kmeans_2	18 919	116 626	6,48%	3,59%
c_kmeans_1	12 707	72 479	4,35%	2,23%
Total	292 141	3 251 229	100,00%	100,00%

Ici, c'est plus intéressant, le plus petit groupe comprend quand même 2,23% des montants. Et les arrêt-ligne-direction-jour sont assez bien répartis dans les différents groupes, il y a 5 groupes avec chacun plus de 10% des arrêt-ligne-direction-jour.

Dès lors, nous allons donc retenir 9 groupes comme nombre adéquat de groupes en faisant un compromis entre la taille des groupes, la définition et la lisibilité que l'on veut.

3.1.3. Identification des neuf groupes d'arrêt-ligne-direction

En observant la Figure 3.7, on se rend compte que la plupart des groupes correspondent à des arrêt-ligne-direction qui ont un seul pic du nombre de montants à une période donnée de la journée. Par exemple, le groupe 6 voit plus de 70% de ses montants entre 15h et 16h.

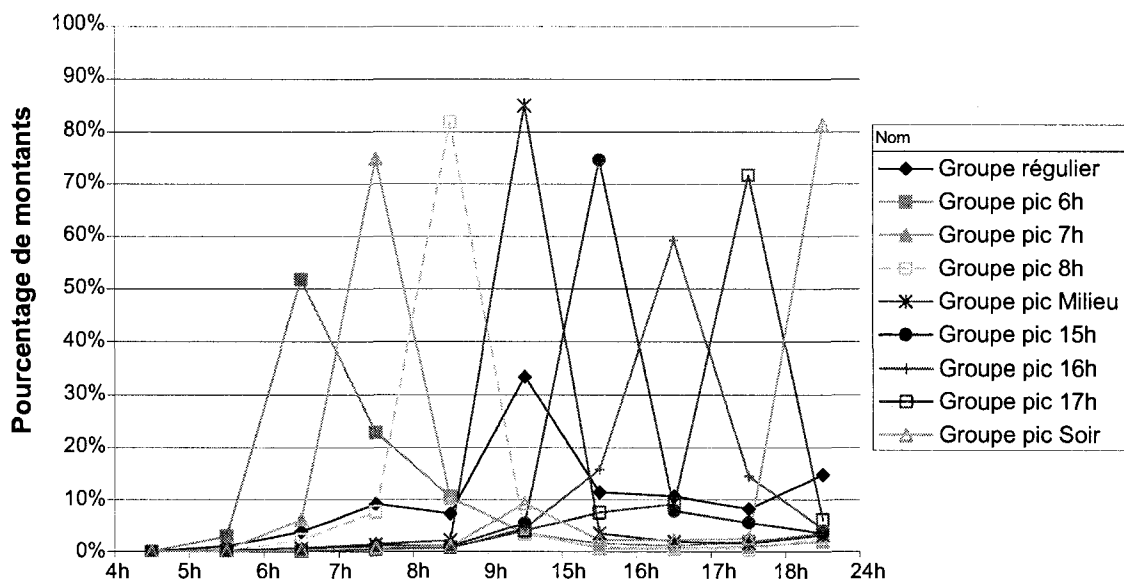


Figure 3.7. Profil de la distribution des montants en fonction de la période de la journée pour les 9 clusters identifiés

Seul le groupe 4 n'a aucune période qui représente plus de 50% du nombre de montants de la journée. Ce groupe correspond à des arrêts qui sont utilisés assez régulièrement entre 7h du matin et la fin de la journée. On observe bien un pic en milieu de journée mais il est seulement d'environ 35% des montants. Pour mieux s'y retrouver nous allons renommer les groupes avec des noms caractéristiques. Ainsi, le groupe 4 va maintenant s'appeler « groupe régulier », les autres groupes vont recevoir le nom de l'heure de leur pic d'activité : le groupe 3 est renommé « groupe pic 6h », le 8 « groupe pic 7h », le 5 « groupe pic 8h », le 7 « groupe pic Milieu », le 6 « groupe pic 15h », le 9 « groupe pic 16h », le 1 « groupe pic 17h », le 2 « groupe pic Soir ».

Dans le SIG de la STO, on ajoute dans la couche de regroupement des données en grille le nombre d'arrêts et le nombre de montants sur les arrêt-ligne-direction pour chaque zone de la grille.

Maintenant, regardons le nombre de montants par heure de la journée sur chaque groupe d'arrêt-ligne-direction.

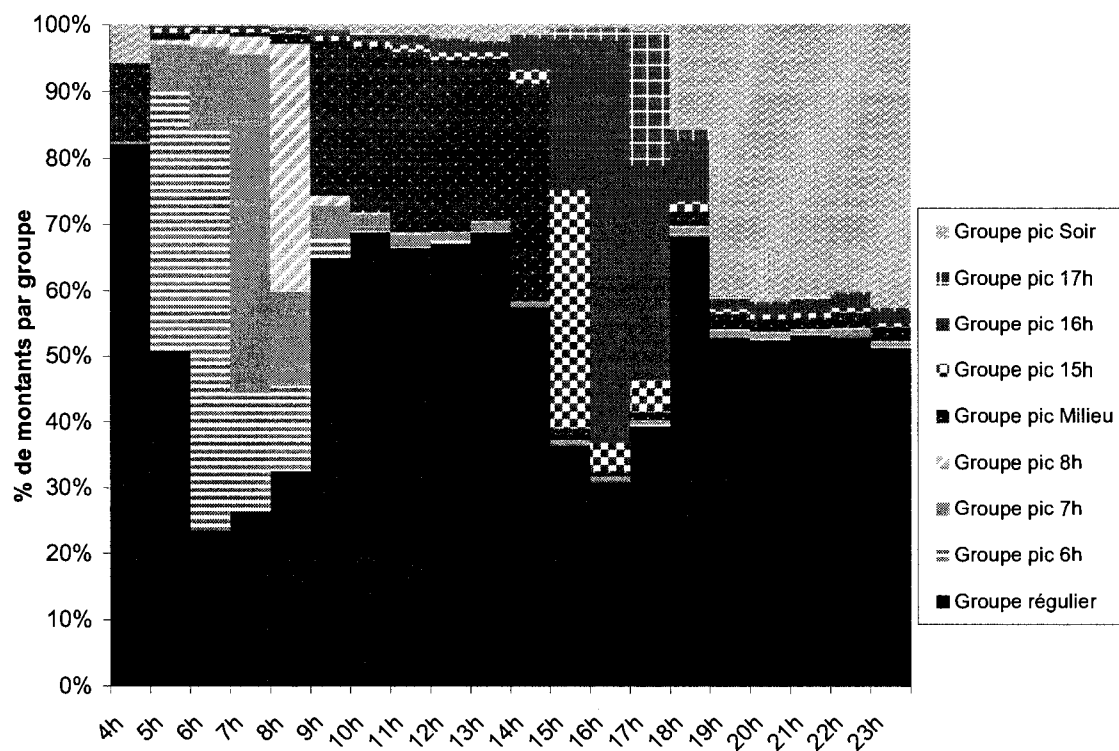


Figure 3.8. Pourcentage de montants par groupe et par heure de la journée

La Figure 3.8 se lit ainsi : 82% des montées enregistrées entre 4h et 5h du matin se font sur des arrêt-ligne-direction du groupe Régulier, 10% sur des arrêt-ligne-direction du groupe Pic Milieu et 7% sur des arrêt-ligne-direction du groupe Pic soir.

On voit (Figure 3.8) qu'en dehors des périodes de pointe, plus de la moitié des usagers montent sur des arrêt-ligne-direction du groupe régulier. Pendant les heures de pointe, à une heure donnée, les usagers utilisent en majorité des arrêt-ligne-direction dont le groupe a un pic pendant cette heure-ci.

3.1.4. Comparaison des segmentations en trois et neuf groupes

Après avoir renommé les groupes de la segmentation en 3 agrégats de la même façon que pour la segmentation en 9, on peut comparer l'appartenance des différents arrêt-ligne-direction-jour aux différents groupes pour les segmentations en 3 et en 9 agrégats. La matrice suivante (Tableau 3.9) présente le nombre d'arrêt-ligne-direction-jour appartenant à chaque cluster de la segmentation en 3 groupes en fonction de l'appartenance dans la segmentation en 9 groupes.

Tableau 3.9. Matrice d'interpénétration des segmentations en 3 et 9 groupes

		Segmentation 3 Groupes			Total
		Groupe Régulier	Groupe pic Matin	Groupe pic Milieu	
Segmentation 9 Groupes	Groupe régulier	33 091	4 813	24 171	62 075
	Groupe pic 6h	20 005	12 099	7	32 111
	Groupe pic 7h	33	52 601	0	52 634
	Groupe pic 8h	24 497	644	0	25 141
	Groupe pic Milieu	0	0	41 300	41 300
	Groupe pic 15h	16 147	195	50	16 392
	Groupe pic 16h	30 674	184	4	30 862
	Groupe pic 17h	12 700	1	6	12 707
	Groupe pic Soir	18 915	4	0	18 919
	Total	156 062	70 541	65 538	292 141

Le Tableau 3.9 se lit ainsi : 33 091 arrêt-ligne-direction-jour du groupe régulier pour la segmentation en 3 groupes sont dans le groupe régulier pour la segmentation en 9 clusters, 4 813 arrêt-ligne-direction-jour du groupe pic matin pour la segmentation en 3 groupes sont dans le groupe régulier pour la segmentation en 9 clusters, etc.

Si on lit le tableau par ligne, on voit que les groupes sont assez stables, les groupes de la segmentation en 9 sont donc pratiquement tous complètement inclus dans un groupe donné de la segmentation en 3. Seuls les groupes réguliers et pic 6h de la segmentation en 9 sont plus étalés. Le groupe régulier de la segmentation en 9 comprend des arrêt-ligne-direction-jour à la fois du groupe régulier mais aussi beaucoup du groupe pic milieu de la segmentation en 3. Le groupe pic 6h contient des arrêt-ligne-direction-jour du groupe régulier et du groupe matin de la segmentation en 3.

Si on lit le tableau en colonne, on voit que le groupe régulier de la segmentation en 3 est assez instable par rapport à la segmentation en 9. Il contient des éléments de tous les groupes de la segmentation en 9 sauf des groupes pic 7h et pic milieu. On peut aussi constater qu'il y a une forte corrélation entre le groupe pic 7h de la segmentation en 9 et le groupe pic Matin de la segmentation en 3 d'une part et entre le groupe pic milieu de la segmentation en 9 et le groupe pic Milieu de la segmentation en 3 d'autre part.

Ces deux groupes ont donc tendance à peu bouger si on augmente le nombre de groupes. Cela veut dire que les comportements des arrêt-ligne-direction-jour dans ces groupes sont très éloignés des comportements des autres éléments et que les éléments à l'intérieur de ces groupes sont assez proches les uns des autres.

3.2. Étude de la variabilité au cours des 4 mois d'observation

3.2.1. Variabilité des groupes d'arrêt-ligne-direction

3.2.1.1. Variation par semaine

L'appartenance des arrêt-ligne-direction aux différents groupes pendant les 4 mois d'observation est présentée dans la figure suivante. On constate (voir Figure 3.9) que le pourcentage d'arrêt-ligne-direction dans chacun des 9 groupes est assez stable sur toute la période d'observation, sauf pour la semaine qui commence le 25 décembre qui correspond à la dernière semaine du mois de décembre où beaucoup d'utilisateurs sont en vacances. Pendant cette dernière semaine de l'année, les groupes régulier et pic Milieu ont plus d'importance que les autres semaines.

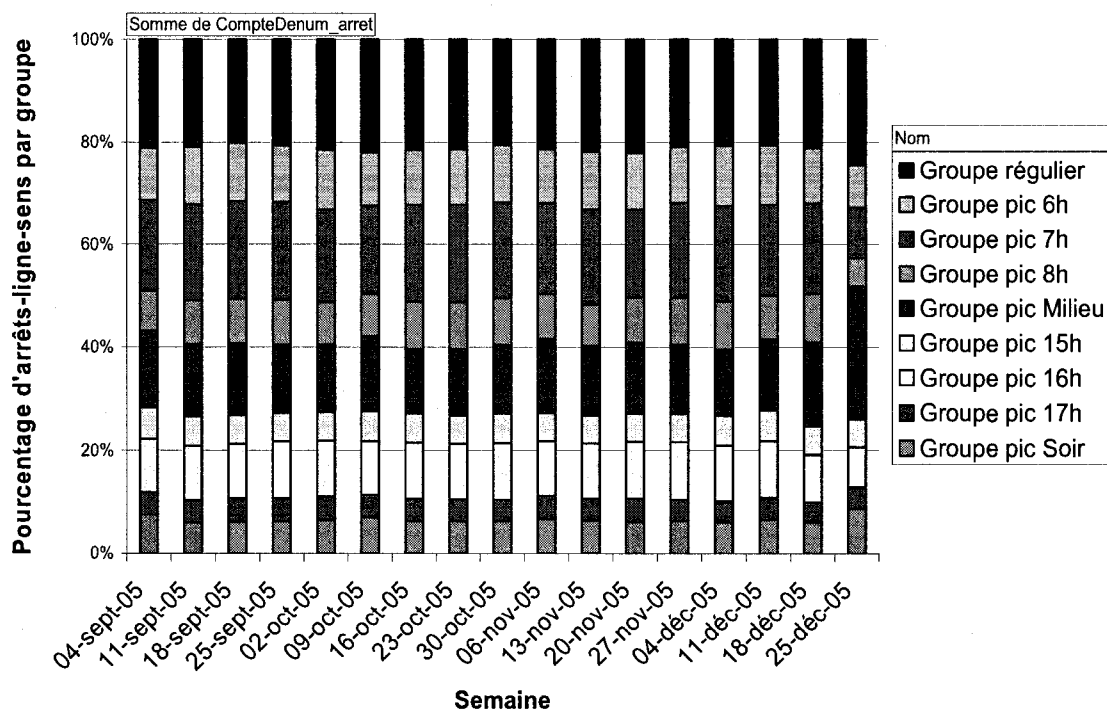


Figure 3.9. Variabilité de l'appartenance des arrêt-ligne-direction aux différents groupes par semaine pendant la période d'observation.

3.2.1.2. Variabilité dans les groupes

Nous allons maintenant étudier la variabilité des arrêt-ligne-direction à l'intérieur des groupes. Il se peut qu'un jour un arrêt ligne-direction soit dans un groupe et qu'un autre jour ce même arrêt-ligne-direction soit dans un autre groupe.

Tout d'abord, nous calculons le nombre de fois qu'un arrêt-ligne-direction est dans un groupe donné sur toute la période d'observation et pour tous les groupes. Ensuite, on peut en déduire le *pourcentage d'appartenance d'un arrêt-ligne-direction à un groupe*. Si un arrêt est présent 10 fois dans le groupe 1, 20 dans le groupe 2 et 20 dans le groupe 3, son pourcentage d'appartenance au groupe 1 est de 20%, au groupe 2 et 3, 40%. Enfin, pour chaque groupe, on calcule le nombre d'arrêt-ligne-direction qui sont toujours dans le même groupe, ceux qui sont entre 75% et 99% des fois dans le même groupe, ceux qui sont entre 50% et 75% des fois, ceux entre 25% et 50% et enfin ceux entre 0% et 25%. On obtient les diagrammes suivants (Figure 3.10):

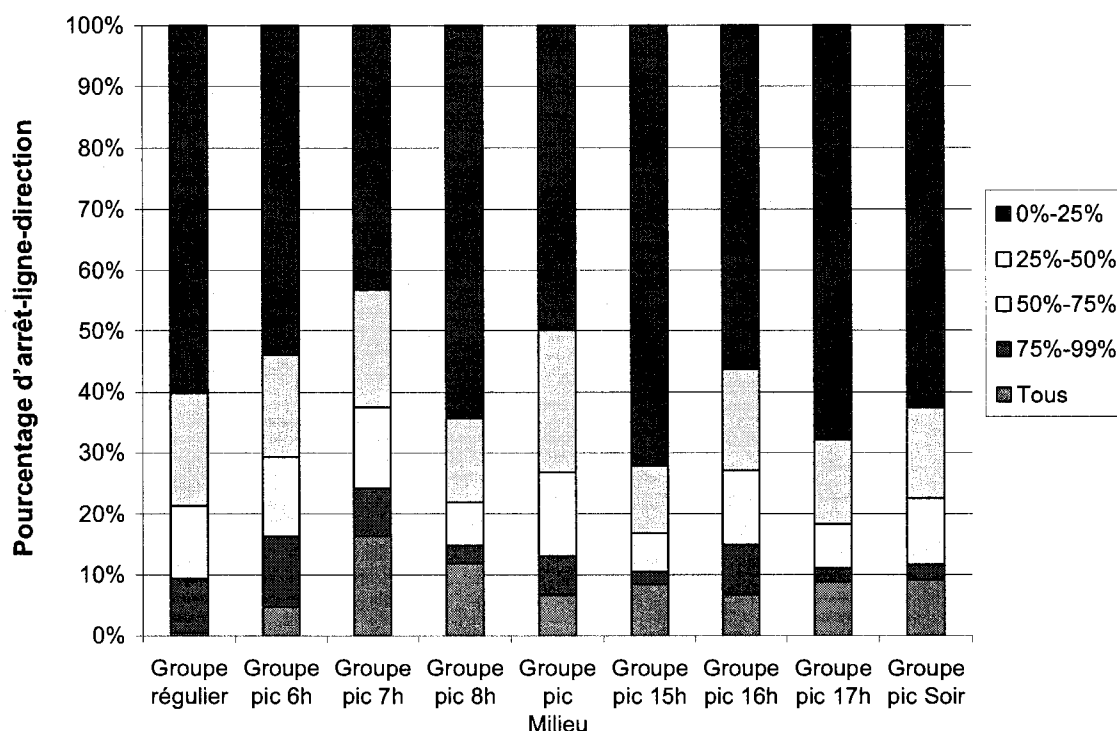


Figure 3.10. Variabilité de l'appartenance des arrêt-ligne-direction

La Figure 3.10 se lit ainsi, dans le groupe pic 6h, 5% des arrêt-ligne-direction sont tous les jours dans ce groupe, 12% sont entre 75% et 99% des jours dans le groupe, 13% sont entre 50% et 75% des jours dans le groupe, 17% sont entre 25% et 50% des jours dans le groupe et enfin plus de la moitié des arrêt-ligne-direction du groupe sont moins de 25% des jours dans le groupe.

On constate une grande variabilité des appartenances aux différents groupes sur la période donnée. Au premier abord, on peut penser que cette variabilité est due à l'influence du jour de la semaine, on est en droit de se demander si un arrêt-ligne-direction a tendance à avoir toujours le même comportement tous les mardis par exemple.

Si on fait les mêmes diagrammes pour chacun des jours de la semaine, la variabilité est bien plus faible un jour donné dans la semaine mais est plus forte le week-end pour la plupart des groupes. Les figures suivantes présentent deux exemples : pour le mardi (Figure 3.11) et le samedi (Figure 3.12), les courbes sont assez semblables les autres jours de la semaine : le mardi est caractéristique des jours de la semaine et le samedi du week-end.

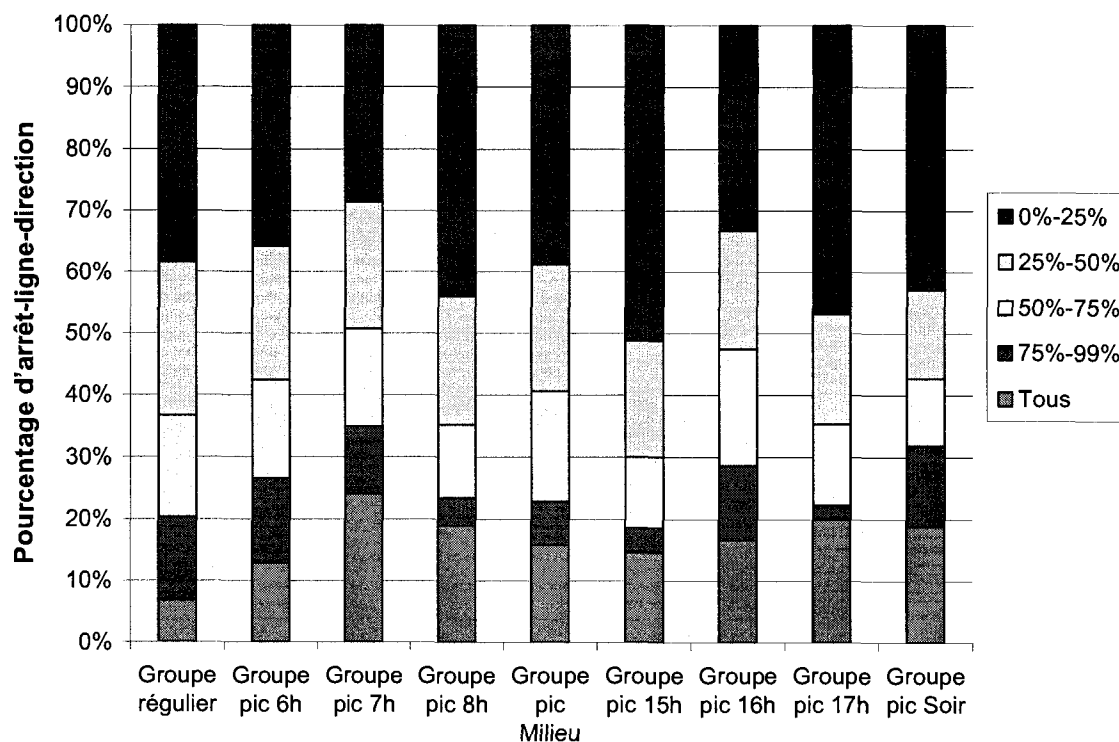


Figure 3.11. Variabilité de l'appartenance des arrêt-ligne-direction au groupe sur tous les mardis

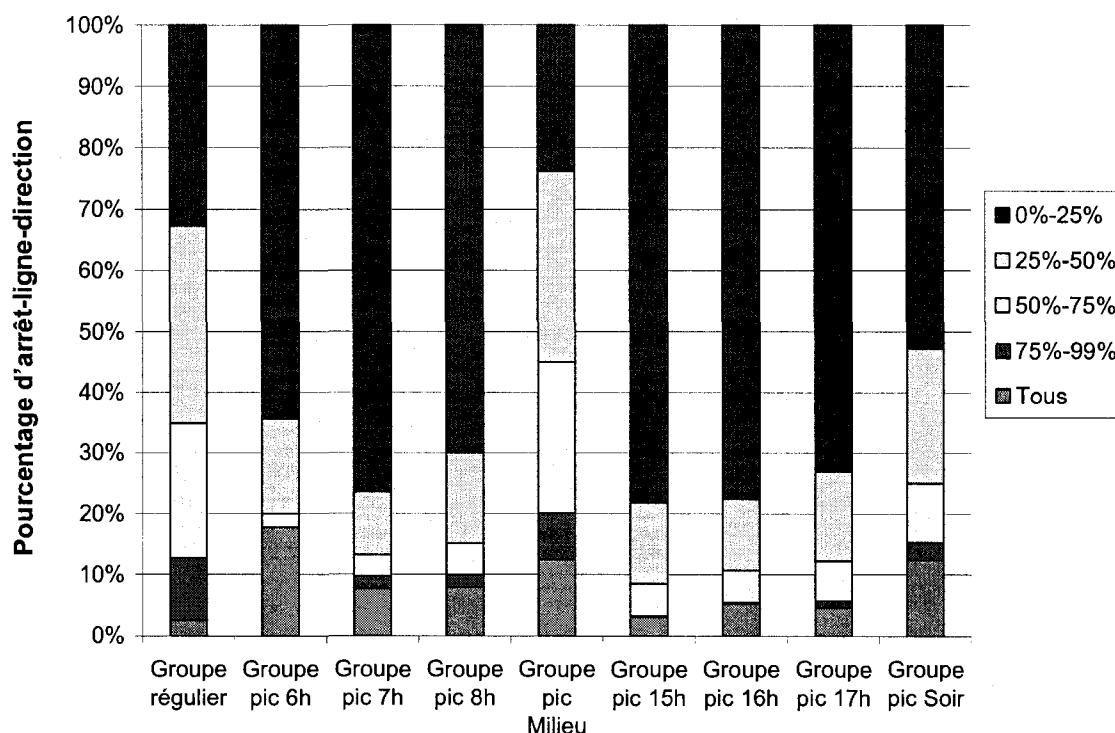


Figure 3.12. Variabilité de l'appartenance des arrêt-ligne-direction sur tous les samedis

Cela veut dire que les arrêt-ligne-direction ne sont pas utilisés tous les jours de la même façon mais que, néanmoins, leur utilisation est plus régulière un jour donné de la semaine.

Malheureusement, cette mesure de la variabilité n'est pas très pertinente. En effet, elle met beaucoup en valeur les arrêt-ligne-direction les plus variables. Ainsi, si un arrêt-ligne-direction a de faibles pourcentages d'appartenance dans tous les groupes, il va être considéré dans chaque groupe. Si un arrêt-ligne-direction est présent dans 10 groupes mais que son taux d'appartenance caractéristique à un seul groupe est élevé, ce que l'on veut savoir c'est qu'il existe un groupe dans lequel il est le plus souvent et pas qu'il existe 9 groupes dans lesquels il est rarement. Dès lors, plutôt que la variabilité dans les groupes, on se propose d'étudier la variabilité entre les groupes.

3.2.1.3. Variabilité entre les groupes

Pour mesurer la variabilité entre les groupes, on regarde pour chaque-arrêt-ligne-direction le groupe dans lequel il est le plus souvent présent (celui où le pourcentage d'appartenance est maximum). On appelle *groupe prépondérant* le groupe ainsi trouvé et *taux caractéristique d'appartenance à un groupe prépondérant* son pourcentage dans ce groupe. Si ce taux est grand cela veut dire que l'arrêt-ligne-direction aura quasiment toujours le même comportement au cours du temps.

Pour calculer ce taux, on construit un tableau qui pour chaque arrêt-ligne-direction, donne le pourcentage d'appartenance de ses arrêts à chaque groupe (cf. Tableau 3.10).

Tableau 3.10. Variabilité de l'appartenance à un groupe des arrêt- ligne-direction

Arrêt	Ligne	Direction	..	Groupe pic 15h	Groupe pic 16h	Groupe pic 17h	Groupe pic Soir
1000	1	0	..	72,38%	0,00%	0,00%	26,29%
1024	1	1	..	0,00%	45,92%	52,50%	0,00%

L'arrêt-ligne-direction 1000-1-0 est 72% des jours dans le groupe pic 15h et 26% des jours dans le groupe pic Soir.

Ensuite, on récupère le taux caractéristique d'appartenance d'un arrêt-ligne-direction à un groupe prépondérant. Ainsi, si on reprend notre exemple, le groupe prépondérant de l'arrêt-ligne-direction 1000-1-0 est le groupe pic 15h avec un taux d'appartenance caractéristique de 72%. Pour cet arrêt-ligne-direction, la dispersion est faible, on peut donc conclure que cet arrêt-ligne-direction a quasiment tous les jours le même comportement. Puis, on construit un graphique avec le pourcentage d'arrêt-ligne-direction qui ont des taux caractéristiques d'appartenance de 100%, d'entre 75% et 99%, d'entre 50% et 75% et d'entre 0% et 25%.

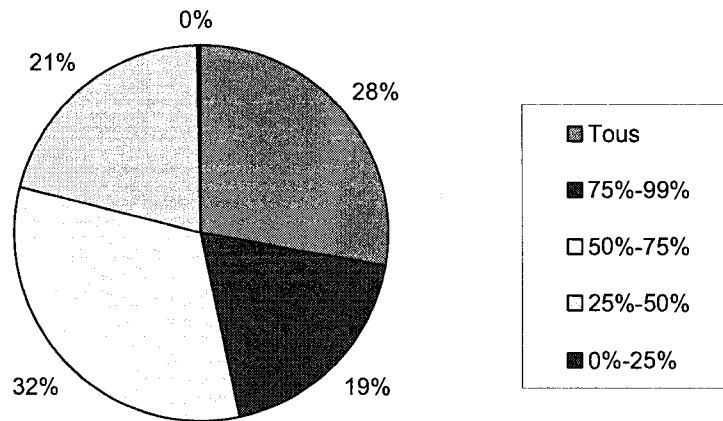


Figure 3.13. Variabilité de l'appartenance des arrêt-ligne-direction à un groupe prépondérant

La Figure 3.13 se lit ainsi : 28% des arrêt-ligne-direction sont toujours dans le même groupe, 19% sont entre 75% et 99% des fois dans le même groupe et 32% sont entre 50% et 75% dans le même groupe, etc.

On voit ici que plus de trois quarts des arrêt-ligne-direction sont au moins 50% des fois dans un même groupe. La variabilité au cours du temps est donc assez faible. On peut donc conclure que, globalement, un arrêt-ligne-direction a quasiment toujours le même comportement au cours du temps. On peut donc assigner un groupe prépondérant à la plupart arrêt-ligne-direction.

On définit maintenant *le noyau d'un groupe* : un arrêt-ligne-direction appartient au noyau d'un groupe si cet arrêt-ligne-direction est plus de 50% des fois dans le groupe, c'est-à-dire si son taux caractéristique d'appartenance à ce groupe est de plus de 50%. Ainsi, si on reprend l'exemple du Tableau 3.10, l'arrêt-ligne-direction 1000-1-0 est dans le noyau du groupe pic 15h. Le noyau correspond donc aux éléments du groupe qui varient peu. Ce sont les arrêt-ligne-direction les plus caractéristiques du groupe qui sont dans son noyau.

Une propriété intéressante du noyau d'un groupe est qu'un arrêt-ligne-direction appartient au plus à un seul noyau de groupe. En effet, il ne peut y avoir qu'un seul groupe qui contient plus de la moitié des occurrences d'un arrêt-ligne-direction. On a vu sur la Figure 3.13 que plus de trois quarts des arrêt-ligne-direction appartiennent au noyau d'un groupe. Le tableau suivant présente le nombre d'éléments dans chaque groupe et dans les noyaux de chaque groupe. Il faut noter que 1375 arrêt-ligne-direction ne sont dans aucun noyau sur les 6537 arrêt-ligne-direction que compte le réseau de la STO. Le Tableau 3.11 présente la répartition des arrêt-ligne-direction (ALD) dans les noyaux de groupes ainsi que la proportion du nombre de montants dans le noyau.

Tableau 3.11. Nombre d'éléments dans le noyau de chaque groupe

	Nb d'ALD différents par groupe	Nombre d'ALD dans le noyau	% d'ALD dans le noyau	Nombre de montants sur les ALD du groupe	Nb montants sur les arrêts du noyau	% montants sur les arrêts du noyau
Groupe régulier	2 407	514	21,35%	1 316 572	1 058 185	80,37%
Groupe pic 6h	1 551	449	28,95%	318 566	228 659	71,78%
Groupe pic 7h	2 789	1 031	36,97%	378 121	227 445	60,15%
Groupe pic 8h	2 521	525	20,83%	147 209	78 883	53,59%
Groupe pic Milieu	2 955	774	26,19%	206 219	74 857	36,30%
Groupe pic 15h	2 472	389	15,74%	202 375	88 387	43,67%
Groupe pic 16h	2 572	667	25,93%	493 062	377 761	76,62%
Groupe pic 17h	2 414	398	16,49%	72 479	14 340	19,79%
Groupe pic Soir	2 027	415	20,47%	116 626	57 117	48,97%
Total	6 537	5 162	78,97%	3 251 229	2 205 634	67,84%

Cela veut donc dire qu'un groupe d'arrêt-ligne-direction est composé d'un noyau d'arrêt-ligne-direction qui ne varie pas beaucoup et d'un nuage d'arrêt-ligne-direction qui varie beaucoup. On peut faire une analogie avec la structure électronique d'un atome (voir Figure 3.14).

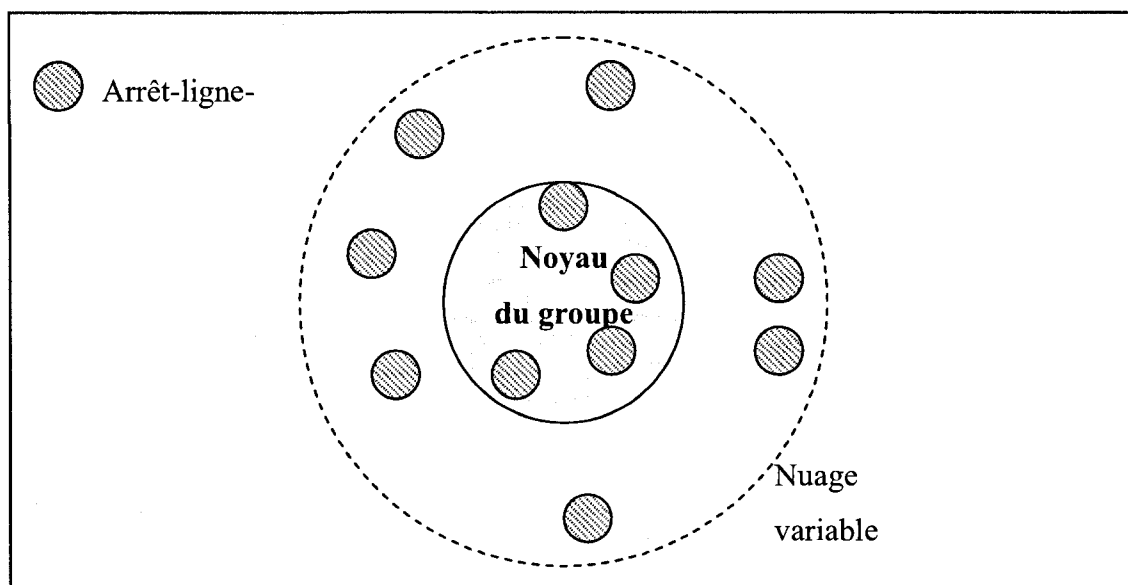


Figure 3.14. Structure « atomique » d'un groupe avec son noyau et son nuage variable

En termes de comportement des usagers, cela signifie que la plupart des usagers ont tendance à prendre le bus très régulièrement sur la plupart des arrêts et qu'un petit nombre d'usagers prend le bus de façon plus irrégulière sur un nombre d'arrêts assez réduit.

On peut maintenant se demander à quel noyau de groupe appartiennent les éléments du nuage variable d'un groupe. Prenons comme exemple le groupe Pic 16h. Ce groupe rassemble 2568 arrêt-ligne-direction différents, et 667 éléments sont dans son noyau. Ensuite, 271 de ses arrêt-ligne-direction sont dans le noyau du groupe régulier, 221 dans celui du groupe pic 17h, 115 dans celui du groupe pic Soir, 111 dans le groupe Pic 15h, 942 n'appartiennent à aucun noyau et un petit nombre d'arrêt-ligne-direction du groupe pic 16h sont dans les noyaux des autres groupes. Si on regarde les profils des montants

de ces différents groupes (Figure 3.7), on voit que les éléments du nuage du groupe Pic 16h sont soit des arrêts qui changent très souvent de groupe soit dans les noyaux de groupe dont le comportement est proche du groupe Pic 16h (par exemple Pic 15h ou Pic 17h).

C'est le cas pour quasiment tous les groupes : la majorité des éléments du nuage d'un groupe appartiennent à des noyaux de groupes dont le comportement est proche du groupe de départ. Cela pourrait être dû au retard des bus (ou des utilisateurs) sur une ligne et à la subdivision de la journée en période d'une heure. Effectivement, si tous les jours, le bus de la ligne 1 dans la direction 0 passe à 8h58 à l'arrêt 1000, il contribuera au pic d'entre 8h et 9h, mais si un jour il a deux minutes de retard et qu'il arrive à 9h, il contribuera au pic suivant. Donc à quelques minutes près, l'arrêt-ligne- pourra ne pas être dans le même groupe mais dans un groupe voisin.

On s'intéresse maintenant à l'influence du jour de la semaine sur la variabilité de l'appartenance d'un arrêt-ligne-direction à un groupe prépondérant. On veut savoir si un arrêt-ligne-direction appartient toujours au même groupe tous les lundis par exemple. On fait les mêmes calculs que précédemment pour chaque jour de la semaine et on obtient l'histogramme suivant (Figure 3.15) :

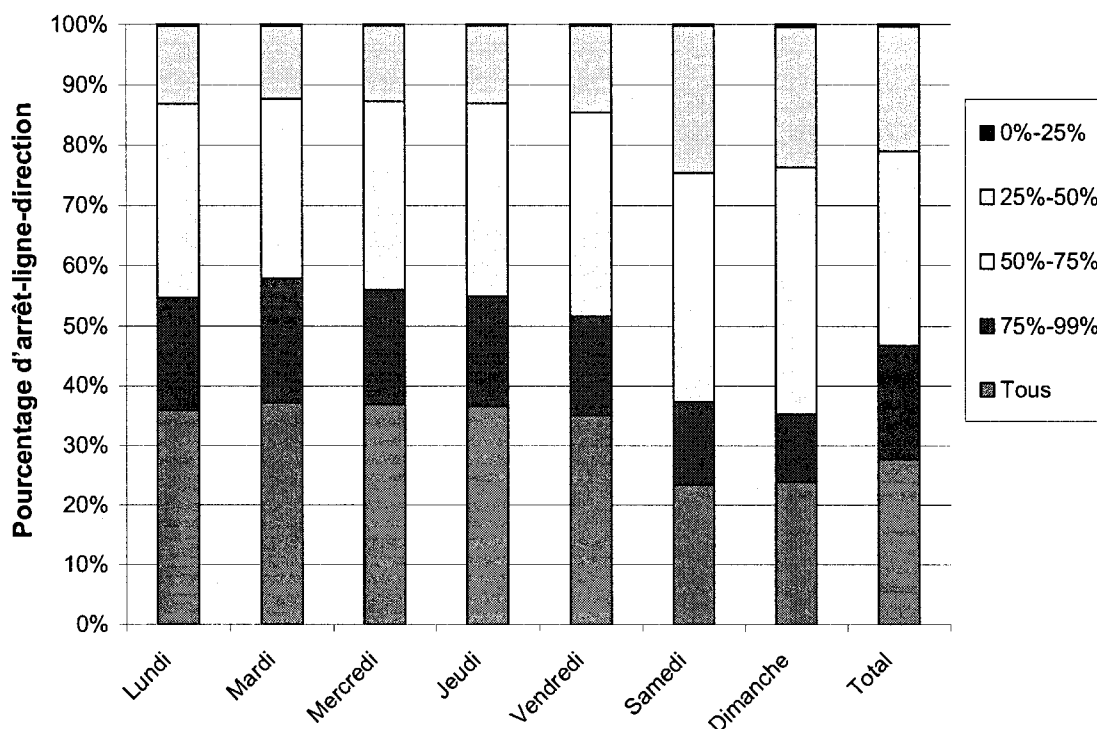


Figure 3.15. Variabilité de l'appartenance des arrêt-ligne-direction à un groupe prépondérant selon le jour de la semaine

La Figure 3.15 se lit ainsi : 36% des arrêt-ligne-direction sont tous les lundis de la période d'observation dans le même groupe, 19% des arrêt-ligne-direction sont entre 75% et 99% des lundis dans le même groupe, et ainsi de suite.

Pendant la semaine (du lundi au vendredi), on voit sur la Figure 3.15 que plus de 85% des arrêt-ligne-direction ont un taux d'appartenance à un groupe prépondérant de plus de 50%. La variabilité sur un jour donné, du lundi au vendredi, est plus faible que la variabilité globale entre les groupes. Cela veut dire que pendant la semaine, les arrêt-ligne-direction auront tendance à avoir un comportement type pour un jour de la semaine mais peuvent avoir un comportement différent un autre jour de la semaine. Du point de vue des usagers, cela signifie qu'ils ont des comportements spatio-temporels très réguliers un jour de la semaine donné du lundi au vendredi (par exemple, ils auront tendance à utiliser le réseau de la même manière, c'est-à-dire aux mêmes endroits et aux

mêmes heures, tous les jeudis). La variabilité augmente en fin de semaine, si un arrêt-ligne-direction appartient à un groupe donné un samedi, on n'est pas certain qu'il appartienne à ce même groupe un autre samedi. Cela confirme encore le fait que l'utilisation du réseau de transport en commun est peu régulière d'un week-end sur l'autre. On peut donc en déduire que les comportements spatio-temporels des usagers le week-end n'ont pas une régularité hebdomadaire.

L'étude de la variabilité dans les groupes a donc permis d'identifier les arrêt-ligne-direction les plus variables, ceux dont le comportement est imprévisible. Ce sont surtout des arrêts qui varient beaucoup à cause du petit nombre d'utilisateurs qui les utilisent. On peut donc assigner à quasiment tous les arrêt-ligne-direction un groupe prépondérant (avec un taux d'appartenance caractéristique de plus de 50%) pour l'ensemble de la période d'étude ou bien pour un jour de la semaine donné. Considérer uniquement le groupe prépondérant permet de diminuer le bruit dans les données. En éliminant cette variation, on a obtenu environ 80% (90% pour un jour de la semaine donné) des arrêt-ligne-direction et 70% des montants qui sont dans un noyau de groupe, les analyses sont donc toujours significatives. On ajoute ce groupe prépondérant dans la couche du SIG des arrêt-ligne-direction.

Cette étude de la variabilité dans les groupes permet aussi de montrer la grande répétitivité des comportements spatio-temporels des usagers pendant la période d'observation. Cette répétitivité est encore plus grande pour un jour donné de la semaine du lundi au vendredi. Cela veut dire que si on connaît la charge du réseau pour un lundi donné par exemple, on peut prévoir assez précisément la charge pour un autre lundi.

3.2.2. Étude de la variabilité le long des ligne-direction

Dans cette section, nous allons voir si les arrêts qui sont sur une même ligne-direction sont utilisés de la même façon, c'est-à-dire si les arrêts d'une ligne-direction sont dans le même groupe. Au total, il y a 211 ligne-direction sur le réseau.

On construit un tableau qui pour chaque ligne-direction donne le pourcentage d'appartenance de ses arrêts à chaque groupe (Tableau 3.12).

Tableau 3.12. Variabilité de l'appartenance à un groupe des arrêts d'une ligne-direction

Ligne	Direction	...	Groupe pic	Groupe pic	Groupe pic	Groupe
			15h	16h	17h	pic Soir
1	0	...	66,40%	0,00%	0,0%	33,60%
1	1	...	0,00%	40,85%	52,50%	0,00%

Les arrêts de la ligne-direction 1-0 sont 66% des fois dans le groupe pic 15h et 33% des fois dans le groupe pic Soir. Ensuite, on récupère le pourcentage d'appartenance maximum de chaque ligne-direction, cette donnée caractérise le groupe prépondérant d'appartenance d'une ligne-direction, c'est le *taux caractéristique d'appartenance d'une ligne-direction à un groupe prépondérant*. Si ce taux est grand cela signifie qu'une grande partie des arrêts de la ligne-direction ont le même comportement. Ainsi, si on reprend notre exemple, le groupe prépondérant de la ligne-direction 1-0 est le groupe pic 15h avec un taux d'appartenance caractéristique de 66%. Pour cette ligne, la dispersion est faible, on peut donc conclure que les arrêts ont quasiment tous le même comportement. Puis, on construit un camembert avec le nombre de ligne-direction qui ont des taux caractéristiques d'appartenance de 100%, d'entre 75% et 99%, d'entre 50% et 75% et d'entre 0% et 25%.

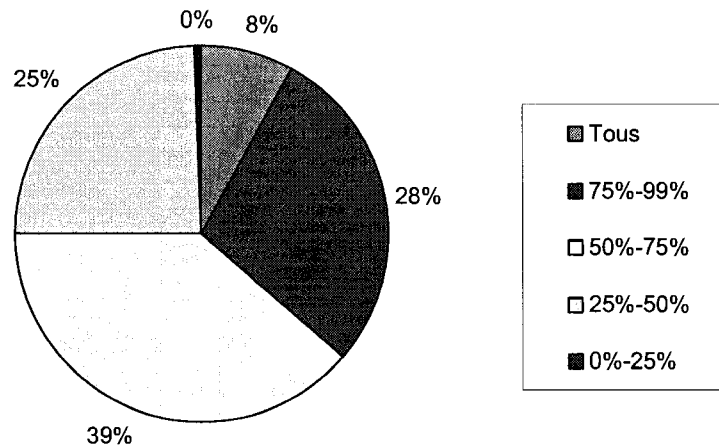


Figure 3.16. Variabilité de l'appartenance des ligne-direction à un groupe prépondérant.

La Figure 3.16 se lit donc ainsi : 8% des ligne-direction ont tous leurs arrêts qui sont toujours dans le même groupe. 28% des ligne-direction ont leurs arrêts qui sont entre 75% et 99% des fois dans le même groupe prépondérant commun à la ligne-direction.

En regardant le graphique obtenu (Figure 3.16), on voit que près de trois quarts des ligne-direction ont un taux caractéristique d'appartenance à un groupe prépondérant assez important (plus de 50%). Ce résultat est assez intéressant, cela veut dire que les arrêts d'une même ligne-direction ont tendance à avoir des comportements homogènes c'est-à-dire les arrêts d'une ligne dans une direction sont utilisés le plus souvent pendant la même tranche horaire de la journée. On pourrait donc affecter à chaque ligne-direction son groupe prépondérant et étudier les groupes de ligne-direction ainsi formés.

On s'intéresse maintenant à l'influence du jour de la semaine sur la variabilité de l'appartenance d'une ligne-direction à un groupe prépondérant. On fait les mêmes calculs que précédemment pour chaque jour de la semaine et on obtient l'histogramme suivant :

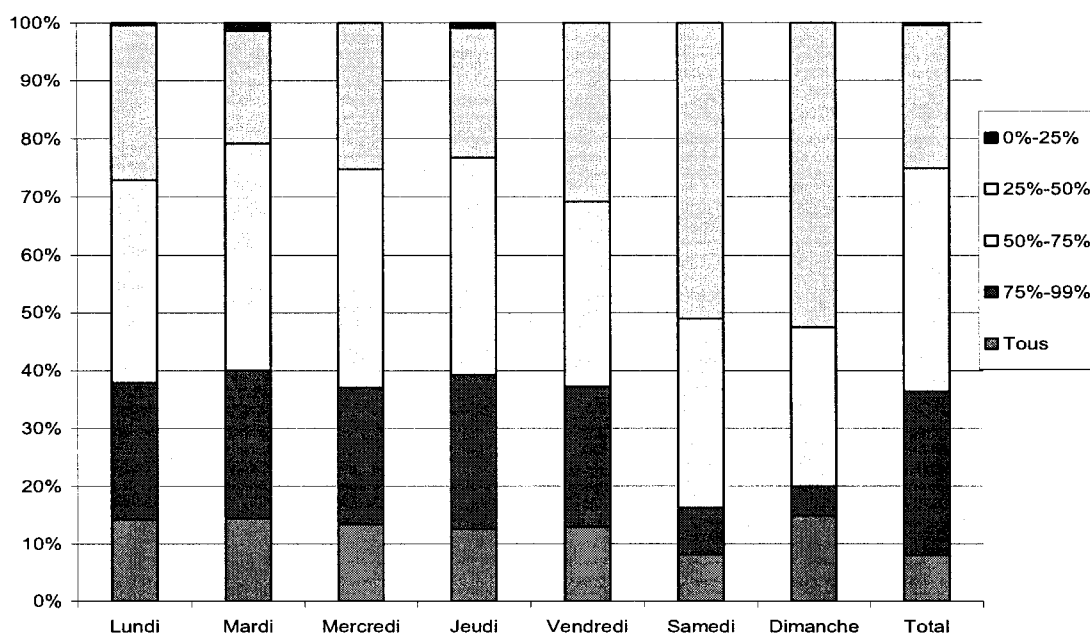


Figure 3.17. Variabilité de l'appartenance des ligne-direction à un groupe prépondérant selon le jour de la semaine.

On constate sur la Figure 3.17 que la variabilité pour les jours du lundi au vendredi est assez faible (plus de 10% des ligne-direction ont un taux d'appartenance caractéristique de 100% du lundi au vendredi). Cela veut dire que pour un jour donné (par exemple le lundi), les ligne-direction ont tendance à avoir toujours le même comportement. Mais la dispersion explose la fin de semaine, plus de 50% des ligne-direction ont un taux d'appartenance caractéristique de moins de 50%. Cela veut dire que les comportements spatio-temporels des usagers sont assez réguliers du lundi au vendredi mais ils deviennent imprévisibles les fins de semaines.

3.2.3. Variabilité des arrêts dans les groupes d'arrêt-ligne-direction

Nous allons voir si un arrêt donné est toujours dans le même groupe quelle que soit la ligne-direction à laquelle il appartient et quel que soit le jour.

On réalise la même étude que dans le paragraphe précédent. Pour chaque arrêt et pour chaque jour de la semaine, on calcule le taux caractéristique d'appartenance à un groupe prépondérant.

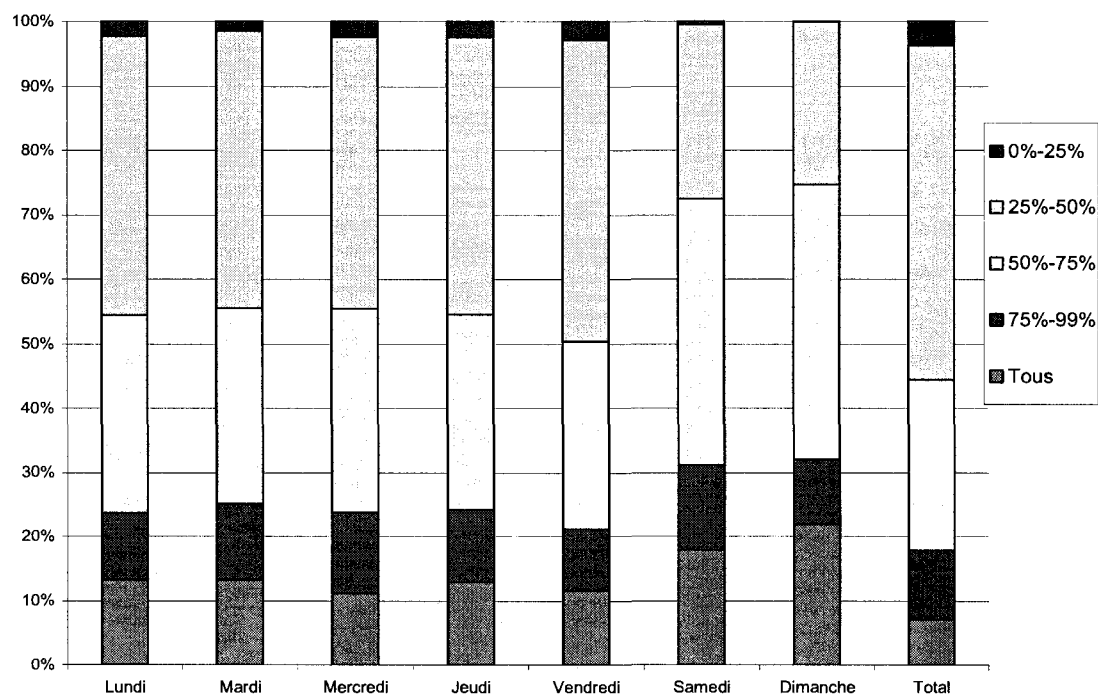


Figure 3.18. Variabilité de l'appartenance des arrêts à un groupe prépondérant selon le jour de la semaine.

Au total, la variabilité d'appartenance des arrêts à un groupe prépondérant est assez grande (voir Figure 3.18). Seulement moins de la moitié des arrêts sont plus de 50% des fois dans le même groupe. Le comportement d'utilisation d'un arrêt par lequel passent plusieurs lignes varie donc en fonction de ces lignes, autrement dit, les lignes qui passent à un arrêt seront utilisées à des heures différentes de la journée. Pendant la semaine, la variabilité des groupes pour un jour donné est aussi assez élevée. Étonnamment, en fin de semaine, la variabilité baisse. Ceci est dû au fait que le nombre d'arrêts utilisés est beaucoup plus faible pendant le week-end. Ici, on ne peut pas vraiment affecter un groupe prépondérant à chaque arrêt.

3.2.4. Influence du jour de la semaine sur les groupes d'arrêt-ligne-direction

Dans cette section, on se propose d'étudier l'influence du jour de la semaine (lundi, mardi,...) sur la répartition des arrêt-ligne-direction dans les différents groupes.

La Figure 3.19 présente la taille de chaque groupe d'arrêt-ligne-direction-jour en fonction du jour de la semaine, les Figure 3.20 et Figure 3.21, le nombre de montants sur chaque groupe d'arrêt-ligne-direction pour chaque jour de la semaine (on a séparé le groupe régulier car il écrasait les autres données).

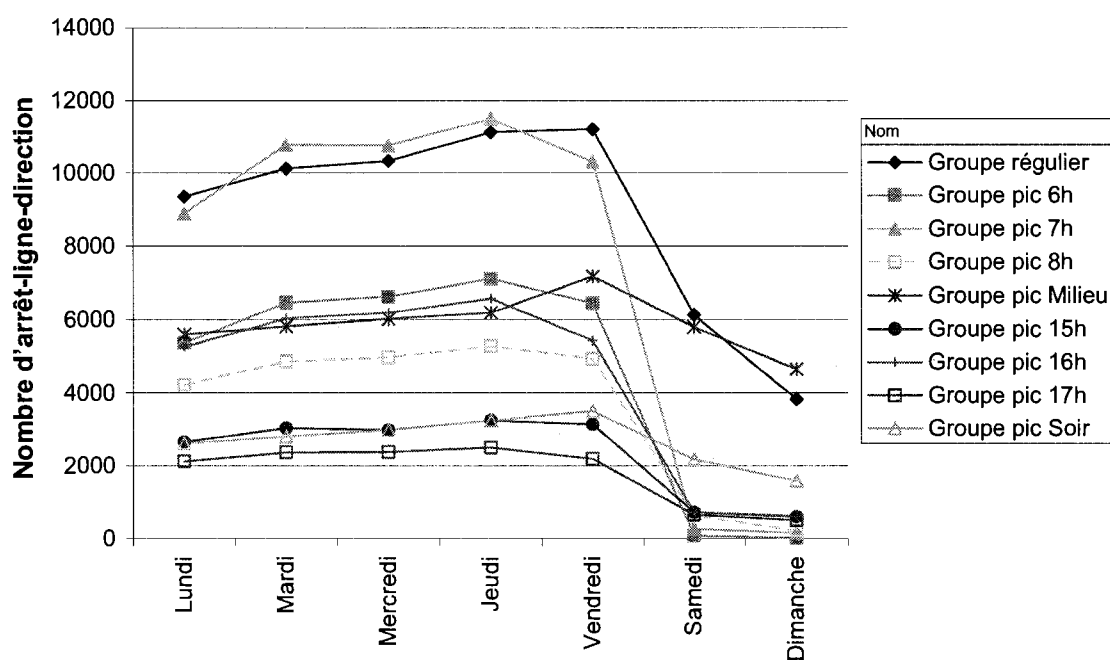


Figure 3.19. Nombre d'arrêt-ligne-direction utilisés par jour de la semaine et par groupe

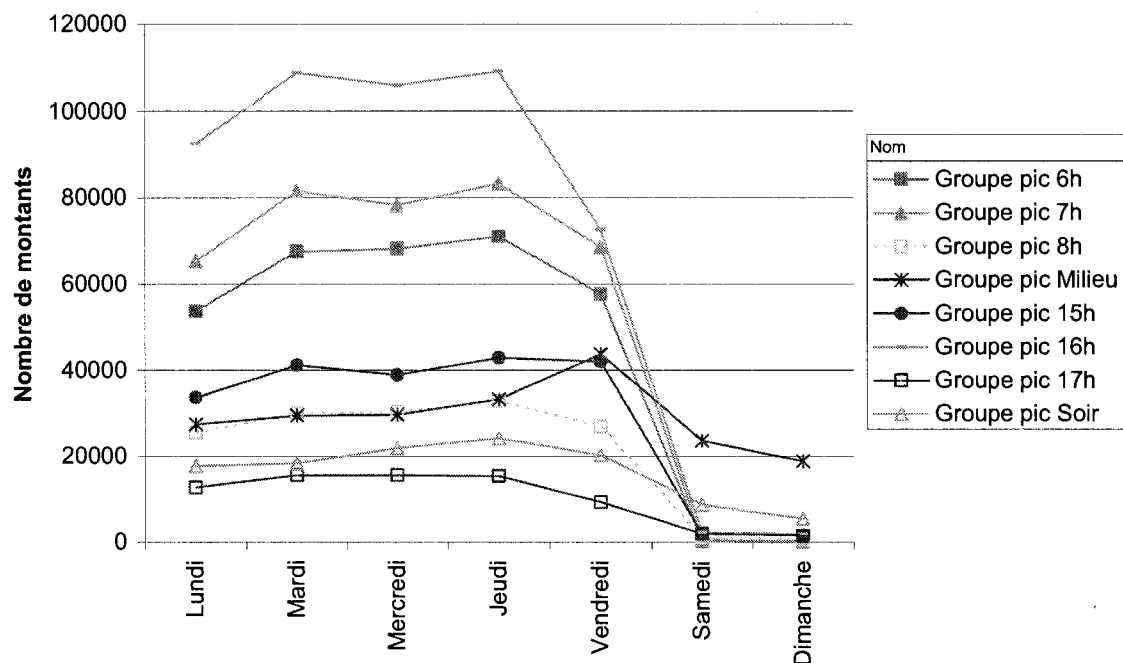


Figure 3.20. Nombre de montants par groupe (autre que régulier) et par jour de la semaine

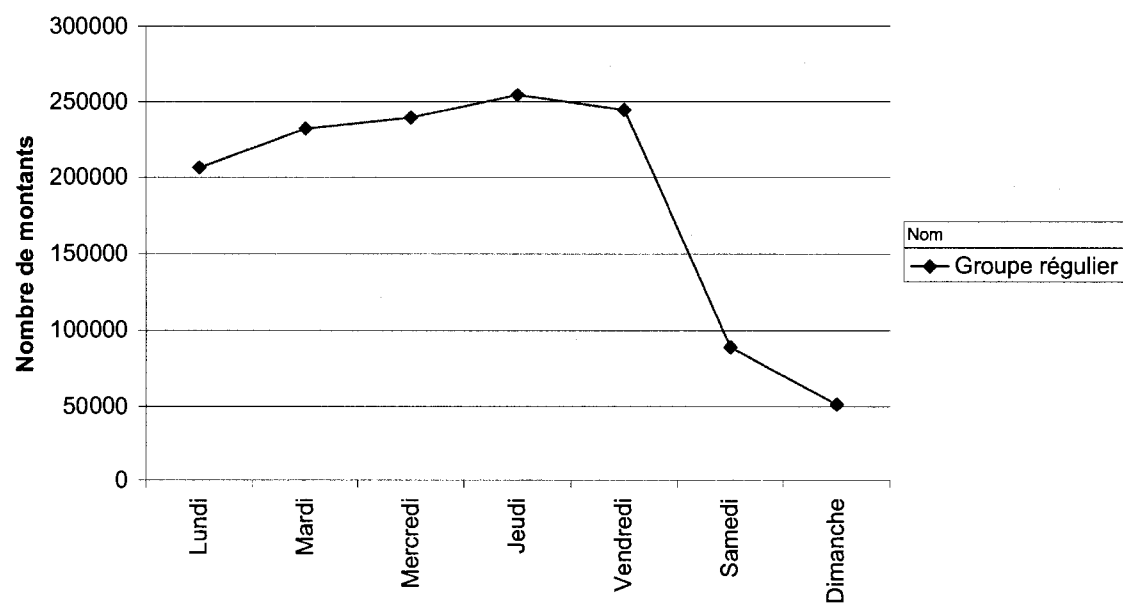


Figure 3.21. Nombre de montants pour le groupe régulier par jour de la semaine

En valeur absolue, on observe (voir Figure 3.19) une forte baisse du nombre d'arrêt-ligne-direction utilisés (ou de montants, la tendance est la même) dans les principaux groupes les fins de semaines, leur nombre étant plus faible le dimanche que le samedi. Ainsi, le groupe pic 6h passe d'une moyenne d'environ 6 000 arrêt-ligne-direction utilisés la semaine à quasiment 0 en fin de semaine.

Maintenant, comparons en valeur relative les contributions de chaque groupe pour chaque jour de la semaine. La Figure 3.22 se lit ainsi : le groupe régulier représente environ 20% des arrêts-ligne-direction du lundi.

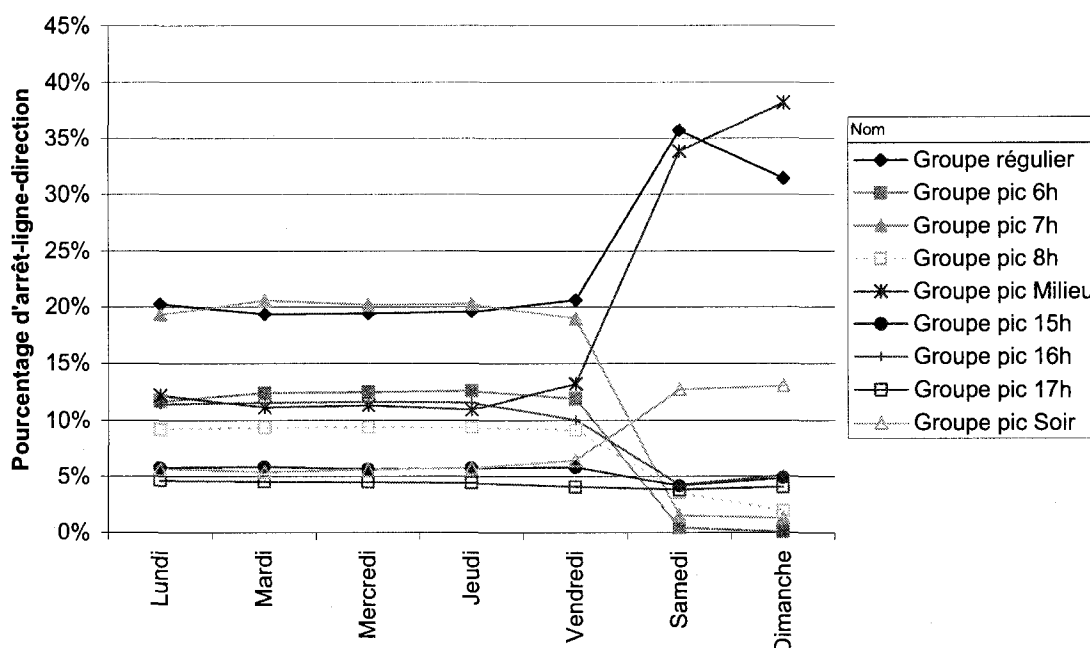


Figure 3.22. Contribution à chaque jour de chaque groupe d'arrêt-ligne-direction

On voit que la proportion des groupes pic 7h, pic 6h, pic 8h et pic 16h chute fortement la fin de semaine. Par exemple, le groupe pic 6h contient environ 6500 arrêt-ligne-direction-jour utilisés la semaine contre 74 le samedi et 7 le dimanche.

Et si on regarde le profil des montants sur ces groupes (voir Figure 3.23), on s'aperçoit que ce sont surtout des groupes qui ont un pic d'utilisation le matin ou en fin d'après-midi. On peut l'expliquer par le fait que les usagers utilisent les transports plus tôt la semaine que le week-end et rentrent chez eux plus régulièrement tout au long de la journée le week-end que la semaine.

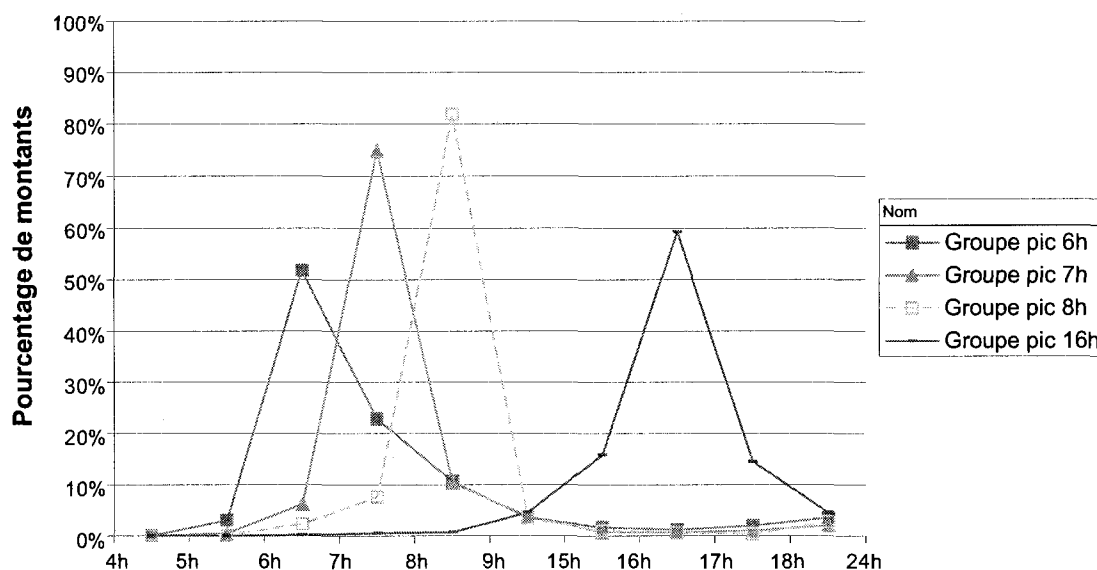


Figure 3.23. Groupes subissant une baisse relative du nombre d'arrêt-ligne-direction le week-end

Les trois principaux groupes qui subissent une forte hausse relative d'utilisation la fin de semaine sont les groupes régulier, pic Milieu et pic Soir (voir Figure 3.24).

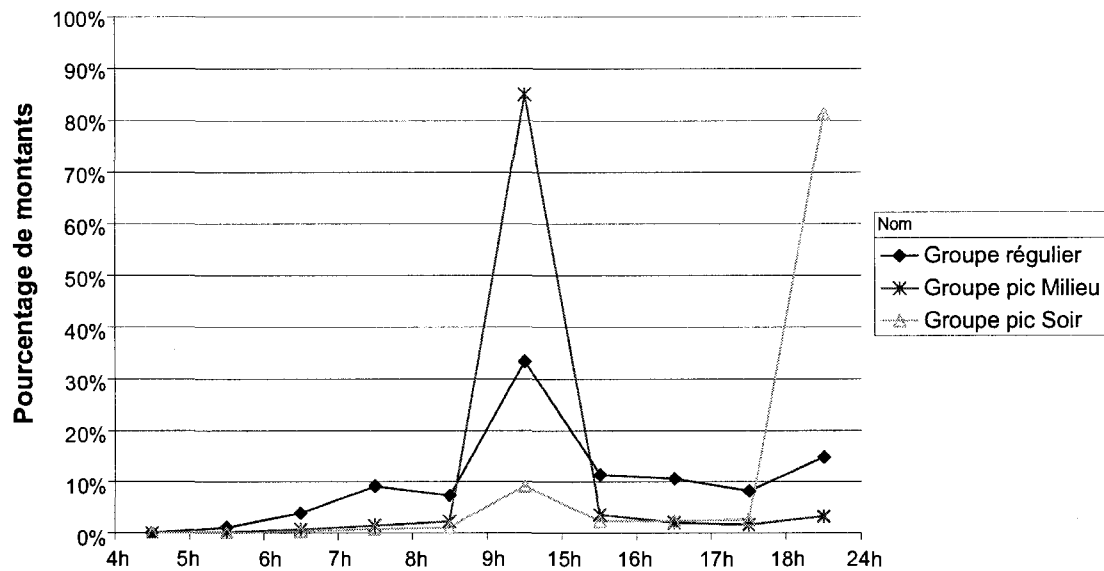


Figure 3.24. Groupes subissant une augmentation relative du nombre d'arrêt-ligne-direction le week-end

On voit que cela correspond surtout aux arrêt-ligne-direction qui sont beaucoup utilisés pendant le milieu de la journée. Cela veut dire que les usagers utilisent le réseau surtout entre 9h et 15h le week-end.

3.3. Étude des comportements des usagers dans les groupes d'arrêt-ligne-direction

Dans cette partie, nous allons étudier les groupes pendant le mois de septembre 2005. Effectivement, les temps de traitement avec Access ne nous permettent pas de traiter toute la base. Dans le fichier des données désagrégées du mois de septembre 2005, on associe à chaque section de déplacement le groupe de l'arrêts-lignes-direction de son arrêt-ligne-direction de montée.

3.3.1. Étude des types d'utilisateurs

Dans cette section, nous allons étudier l'influence du type d'utilisateur (étudiants, adultes, aînés) sur les groupes d'arrêt-ligne-direction.

On détermine la proportion des types d'utilisateurs (aîné, adulte, employé de la STO ou étudiant) dans les différents groupes d'arrêt-ligne-direction (Figure 3.25).

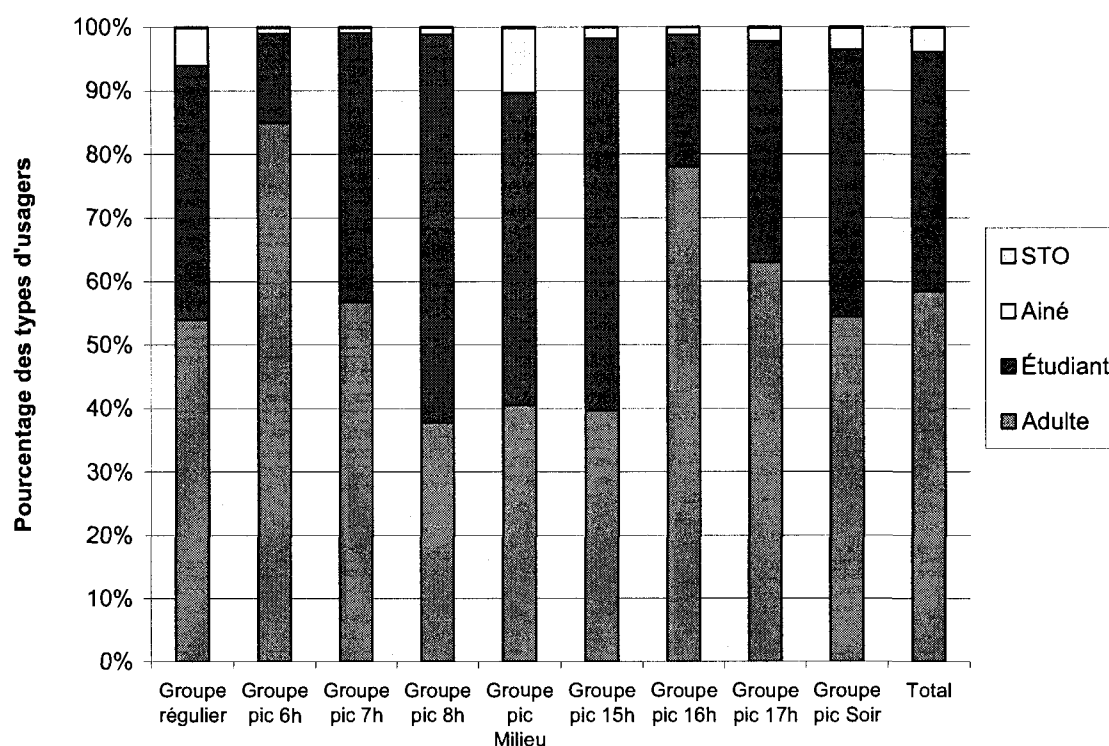


Figure 3.25. Proportion des types d'usagers dans les groupes d'arrêt-ligne

La proportion des types d'utilisateurs du réseau de transport en commun de la STO est très variable d'un groupe à un autre (voir Figure 3.25). On distingue surtout trois types de groupes, les groupes où la proportion est proche de la proportion moyenne, les groupes où il y a beaucoup d'étudiants (groupes pic 8h, pic milieu et pic 15h) et ceux où il y a beaucoup d'adultes (groupes pic 6h et pic 16h). Les aînés utilisent surtout les arrêt-

ligne-direction qui ont un pic d'activité au milieu de la journée et ceux qui sont utilisés régulièrement toute la journée.

Variabilité de la proportion des types d'utilisateur en fonction du jour de la semaine

Du lundi au vendredi, la proportion des types d'usagers dans les différents groupes varie assez peu. Pour les groupes où la proportion des différents types d'usagers est très différente de la moyenne, on observe un resserrement autour de la moyenne les fins de semaine. C'est logique, en fin de semaine, il n'y a plus vraiment de différence dans l'utilisation du réseau entre un étudiant et un adulte.

3.3.2. Origine

On peut représenter sur la carte de l'Outaouais la proportion des montants sur chaque groupe d'arrêt-ligne-direction dans chaque secteur de la grille définie précédemment (voir Figure 2.7). Pour ce faire on utilise le logiciel fGIS. Malheureusement ce logiciel ne permet pas de faire des camemberts avec plus de 8 groupes différents, on réalise donc les camemberts avec seulement les 8 plus gros groupes (en termes de montants), on enlève donc le groupe Pic 17h. Ces 8 groupes représentent plus de 97% des montants. On peut aussi représenter la proportion des arrêt-ligne-direction de chaque groupe et par secteur de la grille (voir Figure 3.27) dans les 8 principaux groupes pendant le mois de septembre 2005.

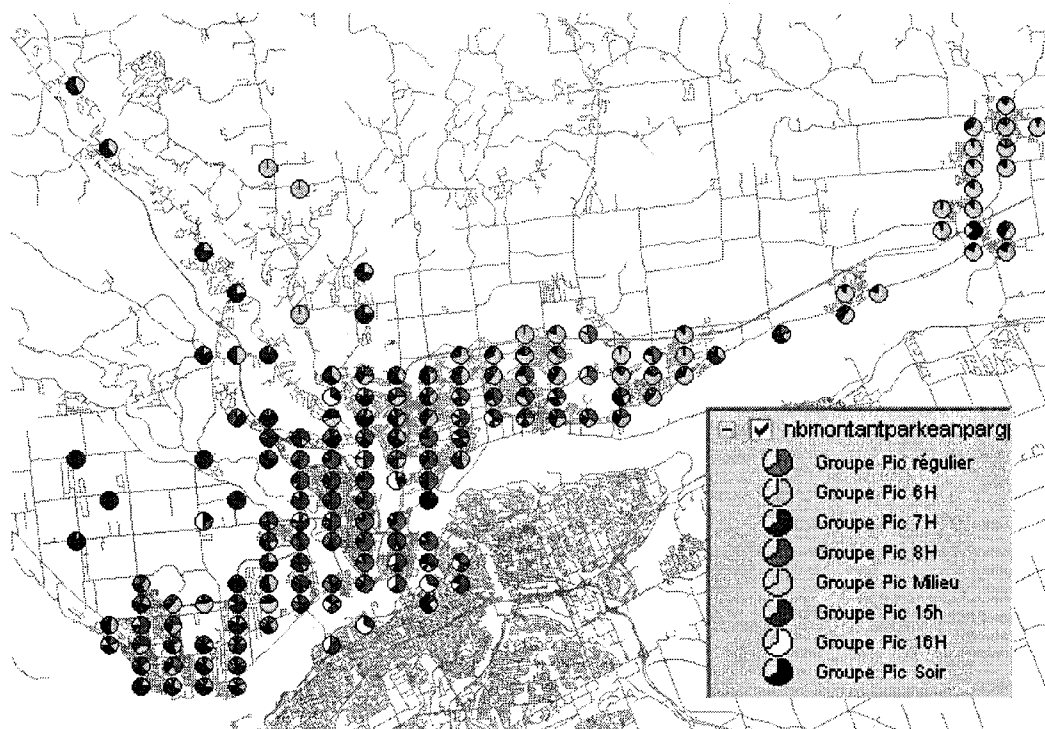


Figure 3.26. Représentation géographique de la proportion de montants dans les principaux groupes d'arrêt-ligne-direction pour le mois de septembre 2005

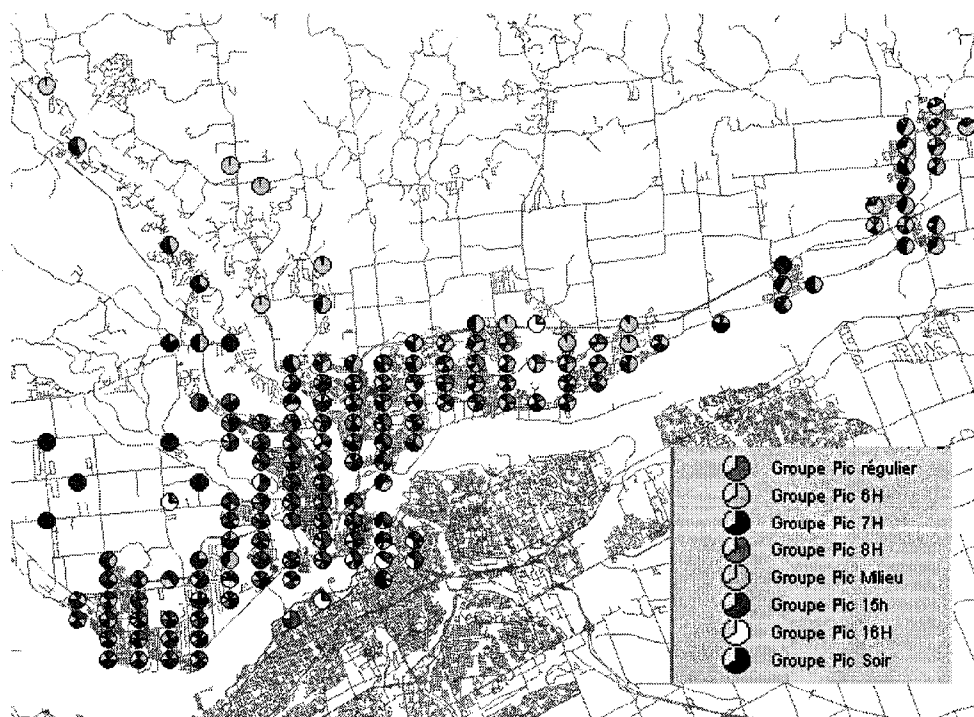


Figure 3.27. Représentation géographique de la proportion d'arrêt-ligne-direction-jour des groupes par secteur

Que l'on prenne la proportion des montants (Figure 3.26) ou des arrêt-ligne-direction-jour (Figure 3.27) dans les différents secteurs de la grille, les tendances principales sont les mêmes. Cette représentation géographique est riche d'enseignements. En effet, on voit que les secteurs les plus éloignés du centre ville sont à majorité en vert clair, c'est-à-dire que les utilisateurs dans ces zones montent en majorité sur des arrêt-ligne-direction du groupe pic 6h. Puis, si on se rapproche du centre ville, on voit la couleur bleu (groupe pic 7h) apparaître. Si on avance encore vers le centre ville, la couleur dominante devient le rouge qui correspond aux arrêts qui sont utilisés régulièrement toute la journée.

On constate donc que plus l'on s'éloigne du centre ville plus les arrêts sont utilisés tôt le matin. C'est assez logique, comme les gens travaillent souvent dans le centre ville et qu'ils commencent à peu près tous à la même heure le matin, ceux qui habitent plus loin doivent partir de chez eux plus tôt pour être à l'heure.

3.3.3. Les distances de déplacement

Dans cette section, nous allons voir si la distance de déplacement dans les différents groupes d'arrêt-ligne-direction est un facteur discriminant. Autrement dit, est-ce que les usagers qui montent sur des arrêt-ligne-direction d'un même groupe ont tendance à faire des déplacements de la même longueur?

Grâce à la connaissance des destinations dérivées, on peut calculer pour chaque trajet (pour lequel une destination a été trouvée) la distance parcourue. Ensuite, on classe ces distances par groupe de pas de 500m et on compte le nombre de sections de déplacement dans chaque groupe de distances. Ainsi, si un usager fait 3200m pendant une section de déplacement, il sera dans le groupe des distances entre 3000m et 3500m. Puis, on calcule la distribution des distances dans chaque groupe d'arrêt-ligne-direction. Enfin, on représente le pourcentage cumulé de la distribution des distances parcourues par section de déplacement dans chaque groupe d'arrêt-ligne-direction sur la figure suivante.

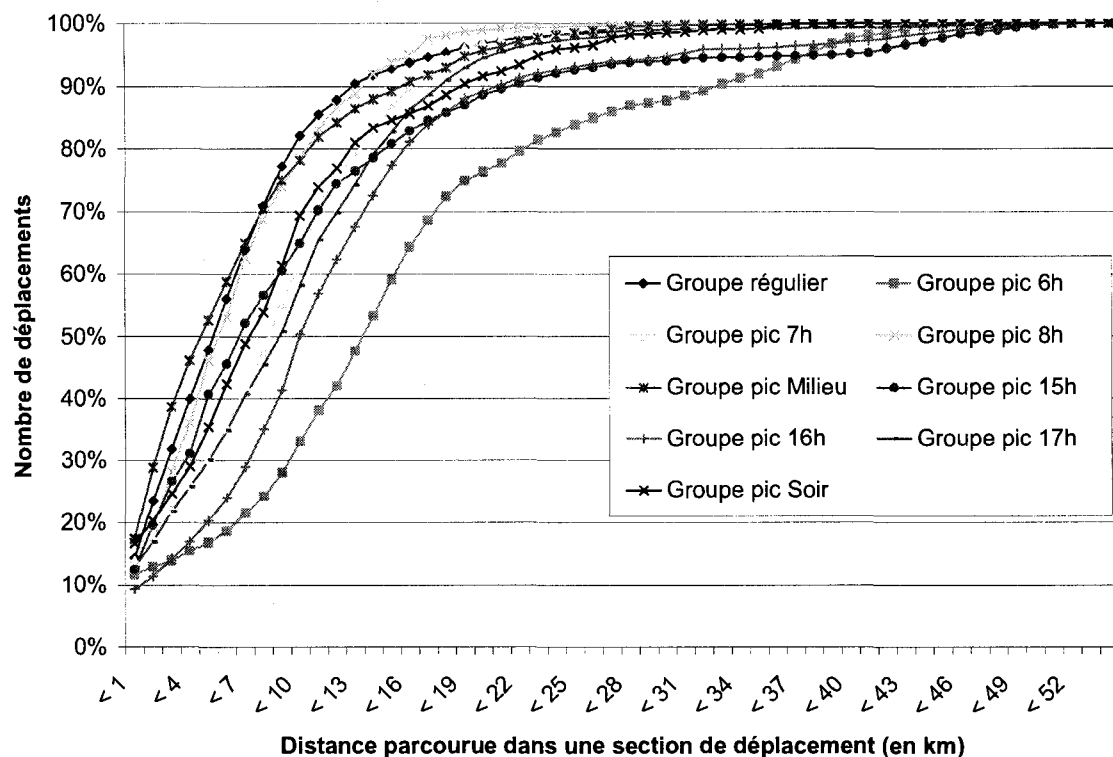


Figure 3.28. Pourcentage cumulé de la distribution des distances parcourues par déplacement dans chaque groupe d'arrêt-ligne-direction

La Figure 3.28 se lit ainsi : 50% des trajets effectués à partir d'un arrêt-ligne-direction du groupe pic 6h font moins de 14km. La pente de la courbe est grande si beaucoup d'utilisateurs font des déplacements situés dans un petit intervalle de distance.

La première conclusion qui saute aux yeux en voyant ce graphique est que la plupart des groupes ont un comportement semblable vis-à-vis de la distance, les sections de déplacement les plus nombreuses sont les plus courtes. Ainsi, pour la plupart des groupes, plus de la moitié des sections de déplacement font moins de 10km et plus de 90% font moins de 20km.

Un seul groupe est vraiment singulier, c'est le groupe pic 6h. La plupart des déplacements des utilisateurs qui montent sur les arrêt-ligne-direction de ce groupe font de

grandes distances. Le groupe pic 6h est le groupe des arrêts-ligne-direction qui sont utilisés le plus tôt. Ceci vient confirmer ce que l'on a déjà dit : les usagers qui utilisent le réseau de transport en commun le plus tôt le font car ils travaillent loin de leur lieu de résidence.

Les groupes qui atteignent le plus vite 50% sont les groupes régulier, pic 8h, pic milieu et pic 15h : ces groupes sont ceux où la plupart des déplacements sont « courts » (la moitié des voyages font moins de 6km). Ces groupes correspondent aux deux catégories suivantes:

- les groupes à majorité étudiante (les groupes pic 8h, pic milieu et pic 15h). Les usagers vont à l'école qui est située près de chez eux
- le groupe régulier qui correspond aux arrêts utilisés toute la journée. Ce sont en majorité des arrêts du centre ville.

3.3.4. Étude des groupes à majorité étudiante

On représente dans chaque zone du découpage en grille, défini au paragraphe 2.4.4, le nombre de montants (symboles rouges) et de descendants (symboles bleus) pendant le mois de septembre 2005 sur les arrêts-ligne-direction des 3 groupes à majorité étudiante (groupe pic 8h (Figure 3.29), groupe pic milieu (Figure 3.30) et groupe pic 15h (Figure 3.31)). Sur ces cartes, la taille des symboles est proportionnelle au nombre d'usagers; cela permet de voir les zones les plus utilisées en termes de montée et de descente.

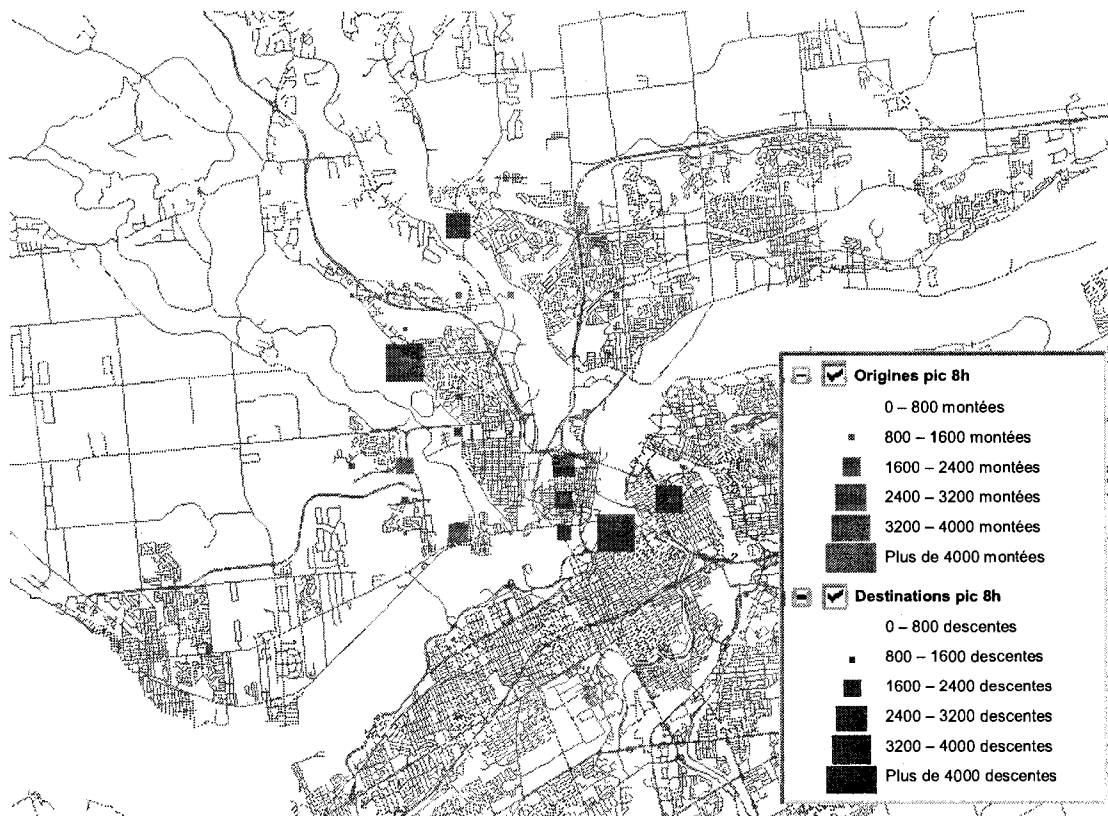


Figure 3.29. Nombre d'utilisateurs par zone d'origine (rouge) et de destination (bleu) du groupe pic 8h pour le mois de septembre 2005

En observant la Figure 3.29, on voit qu'il y a 7 zones principales de destinations pour les utilisateurs qui montent sur des arrêt-ligne-direction du groupe pic 8h. Le matin, on peut penser que les étudiants se rendent à l'école et les secteurs les plus importants en termes de destination correspondent effectivement à des lieux où il y a des écoles. Seulement une zone ne possède pas d'école, c'est le centre ville d'Ottawa qui attire beaucoup d'utilisateurs à toute heure de la journée. Ainsi, les six autres secteurs principaux de la grille contiennent 3 cégeps et 6 écoles secondaires.

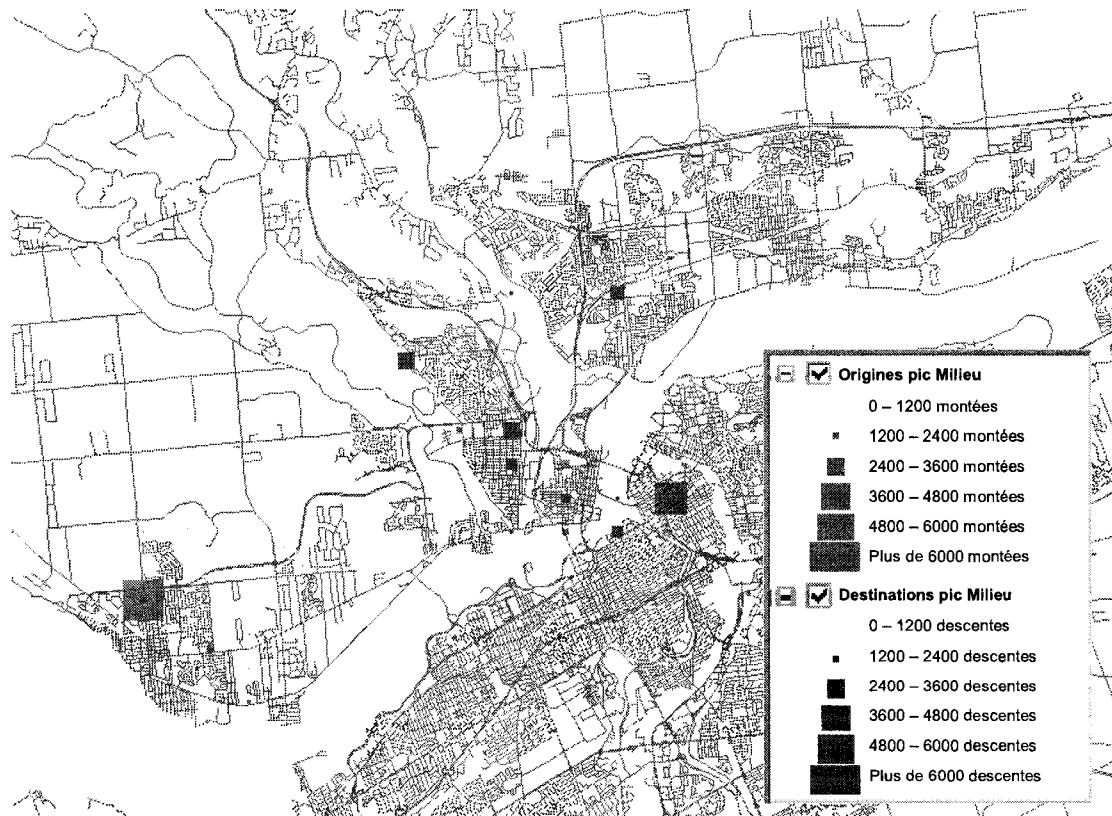


Figure 3.30. Nombre d'utilisateurs par zone d'origine (rouge) et de destination (bleu) du groupe pic Milieu pour le mois de septembre 2005

La Figure 3.30 montre les origines et les destinations des utilisateurs qui montent sur un arrêt-ligne-direction du groupe pic Milieu. Ce groupe rassemble les arrêts qui sont surtout utilisés entre 9h et 15h. Pour les étudiants, ce groupe contient les étudiants qui commencent tard (après 9h) ou bien qui finissent tôt (avant 15h), c'est pour cela que l'on a à la fois des secteurs importants pour les destinations et pour les origines. La zone où l'on observe le plus de montée contient la plus grande école secondaire de l'Outaouais, l'École Secondaire Grande rivière où il y a environ 2 000 élèves. Cela doit vouloir dire que beaucoup d'élèves de cette école finissent les cours avant 15h.

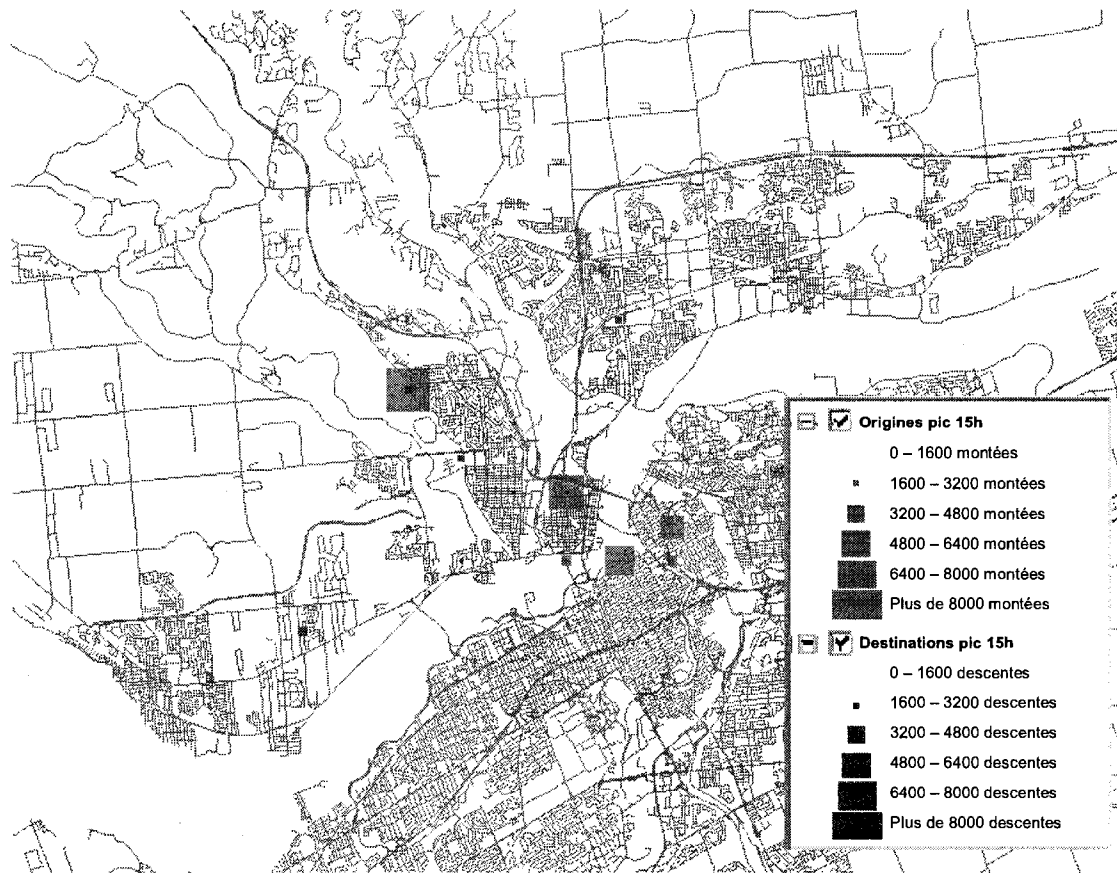


Figure 3.31. Nombre d'utilisateurs par zone d'origine (rouge) et de destination (bleu) du groupe pic 15h pour le mois de septembre 2005

La Figure 3.31 montre les origines et les destinations des utilisateurs qui montent sur un arrêt-ligne-direction du groupe pic 15h. Ce groupe rassemble les arrêts qui sont surtout utilisés entre 15h et 16h, début de la période de pointe de l'après-midi. Ces déplacements sont donc surtout des « retours », ils permettent d'identifier les lieux d'activités de la journée. Les deux zones où l'on observe le plus de montants sur des arrêt-ligne-direction du groupe pic 15h contiennent des écoles :

- la zone la plus importante contient un cégep (Cégep de l'Outaouais, Campus Gabriel Roy : 2 600 étudiants environ) et une école secondaire (école secondaire Mont-Bleu : 1 300 élèves environ)
- la deuxième contient une école secondaire (école secondaire de l'île, 1 300 étudiants environ)

- les deux autres zones les plus importantes en termes d'origine correspondent au centre ville d'Ottawa d'où les travailleurs commencent à partir. Les destinations sont disséminées sur tout le territoire, cela correspond aux domiciles des usagers

En étudiant les origines et les destinations des groupes à majorité étudiante, on peut logiquement identifier les zones dans lesquelles sont situées les écoles les plus importantes de l'Outaouais.

3.4. Influence spatio-temporelle des lieux d'activités sur un réseau de transport en commun

L'objectif global de notre étude est de comprendre l'influence des lieux d'activités sur l'utilisation d'un réseau de transport en commun. Pour ce faire, nous allons, dans un premier temps, mettre en place les relations de proximité entre les différents éléments du SIG. Puis, dans un deuxième temps, nous allons étudier les comportements des usagers au voisinage des lieux d'activités. Et enfin, on se propose de lier les groupes d'arrêt-ligne-direction faits à partir des analyses de data mining à leur environnement géographique, c'est-à-dire aux générateurs de déplacements situés à proximité.

3.4.1. Mise en place des relations de proximité

On définit une relation de proximité entre les arrêt-ligne-direction et les principaux générateurs de déplacements du SIG identifiés en 2.4.5. On dit qu'un arrêt est proche d'un élément ponctuel s'il se trouve à moins de 500m de l'élément en question (distance euclidienne). Pour les nœuds du réseau, on ne considère pas la proximité mais on regarde si l'arrêt-ligne-direction en question est un nœud du réseau. On définit aussi l'opérateur d'inclusion pour les objets surfaciques (un arrêt peut être inclus dans le centre ville).

Maintenant, on regarde pour chaque arrêt-ligne-direction s'il est proche (ou inclus dans) d'un générateur de déplacements. On construit donc un tableau binaire qui pour chaque arrêt-ligne-direction nous dit s'il est proche ou non de chaque générateur de déplacements (

Tableau 3.13). On peut noter que les cégeps, les écoles secondaires et les universités sont regroupés dans la catégorie « école ».

Tableau 3.13. Relation de proximité des arrêt-ligne-direction avec les générateurs de déplacements

Arrêt-ligne-direction	Proche parcobus	Proche école	Proche centre d'achat	Inclus dans centre ville	Proche hôpital	Nœud du réseau
3002-67-0	1	1	1	0	0	1
3000-39-0	1	1	1	0	0	1
2722-67-1	0	0	0	1	0	1
2108-37-1	0	0	1	0	0	1
3003-78-1	1	1	1	0	0	1
2066-37-1	0	1	0	0	1	1
2151-39-1	1	0	0	0	0	1

Par exemple, si on zoome sur une partie du quartier de Hull (Figure 3.32), on s'aperçoit que certains arrêts sont proches de plusieurs générateurs de déplacements. Ceci va tendre à polluer nos analyses, car on ne va pas pouvoir séparer les influences des divers lieux d'activités.

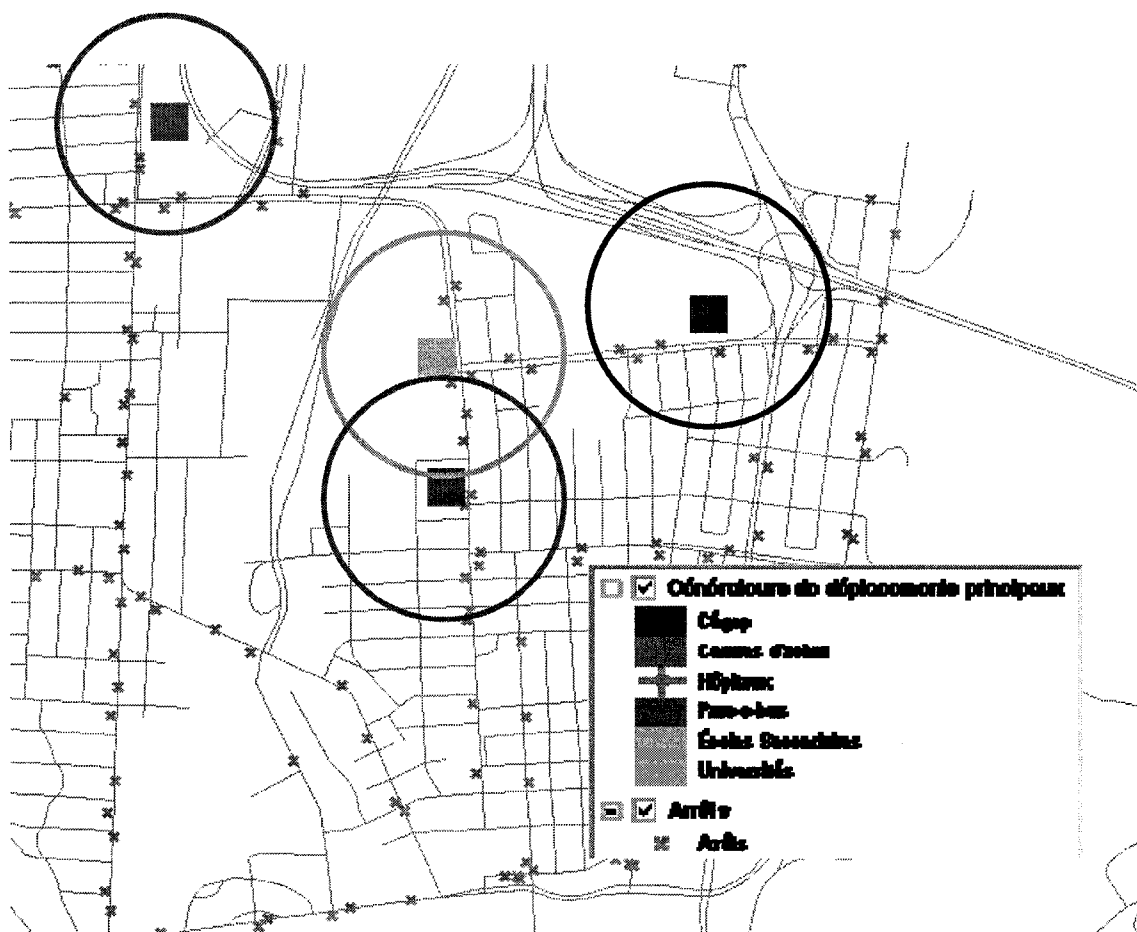


Figure 3.32. Proximité des générateurs de déplacements et des arrêts

On intègre cette nouvelle table dans la base de données désagrégée du mois de septembre 2005.

3.4.2. Études des usagers aux les arrêts proches des lieux d'activités

Dans cette section, nous allons étudier les profils des usagers qui montent aux arrêts qui sont situés à proximité d'un générateur de déplacements donné et nous allons analyser l'influence des lieux d'activités sur les classes d'usagers. Pour ce faire, nous allons prendre comme exemple les arrêts situés à proximité des écoles et ceux situés à proximité des centres commerciaux.

3.4.2.1. Comportements des usagers aux arrêts situés proches des écoles

On calcule la distribution des montants pour chaque période d'une heure de la journée pour tous les jours de la semaine pour tous les arrêts situés à proximités d'une école.

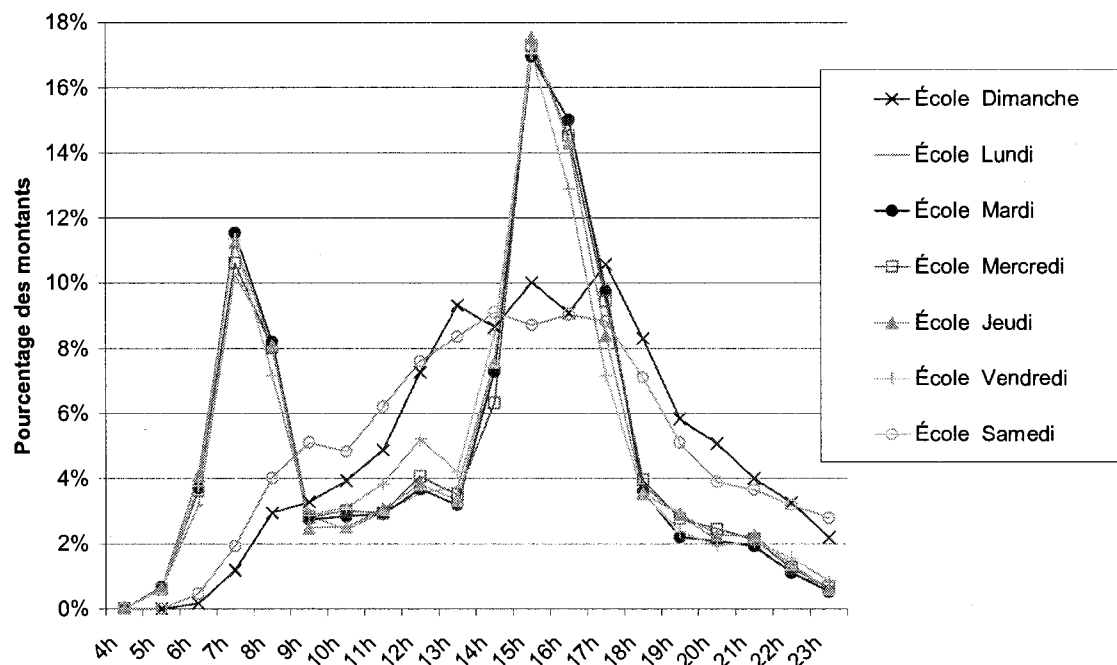


Figure 3.33. Profil des montants aux arrêts proches des écoles par heure et par jour de la semaine

Les comportements des usagers qui montent aux arrêts proches des écoles sont très semblables tous les jours de la semaine du lundi au vendredi, avec un pic d'activité entre 15h et 16h, soit une heure plus tôt que le pic d'utilisation global de l'heure de pointe de l'après-midi (voir Figure 2.11). Toutefois, le vendredi, on observe plus d'utilisation pendant la période du diner, entre 11h et 14h, que les autres jours de la semaine. Le week-end, l'utilisation du réseau est encore une fois très différente du reste de la semaine, en effet, les écoles ne jouent plus leur rôle attracteur de lieux d'activités le samedi et le dimanche.

Maintenant, nous allons voir l'influence des écoles sur les classes d'utilisateurs qui montent à proximité.

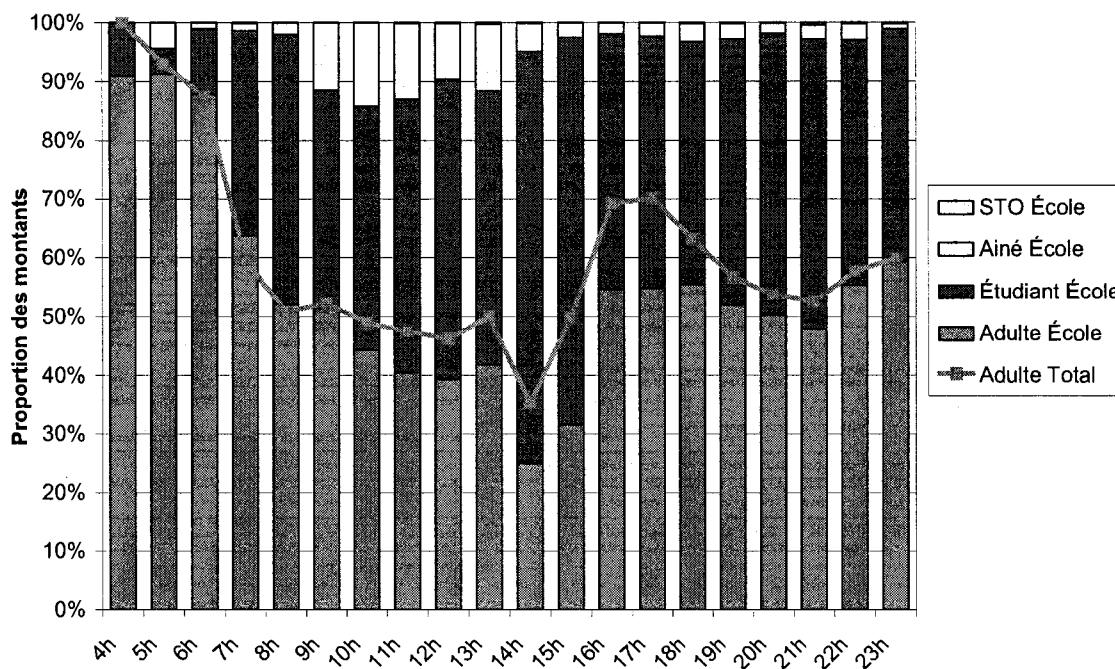


Figure 3.34. Proportion des classes d'utilisateurs montants aux arrêts proches des écoles pour tous les jeudis de septembre 2005 par heure de la journée

La figure précédente (Figure 3.34) se lit ainsi :

- la courbe « Adulte Total » représente le pourcentage moyen de montants de classe adulte sur tous les arrêts de la STO pendant les jeudis de septembre 2005 (voir 2.5.5) par heure
- l'histogramme montre les proportions moyennes de chaque classe d'utilisateurs qui montent sur des arrêts situés à proximité des écoles pour tous les jeudis de septembre 2005 par heure de la journée

On se rend compte qu'à partir de 11h et jusqu'à 19h, la proportion d'étudiants montants aux arrêts proches des écoles est très supérieure à la proportion totale. C'est tout à fait normal que les étudiants qui soient majoritaires aux arrêts proches des écoles.

3.4.2.2. Comportements des usagers aux arrêts situés proches des centres commerciaux

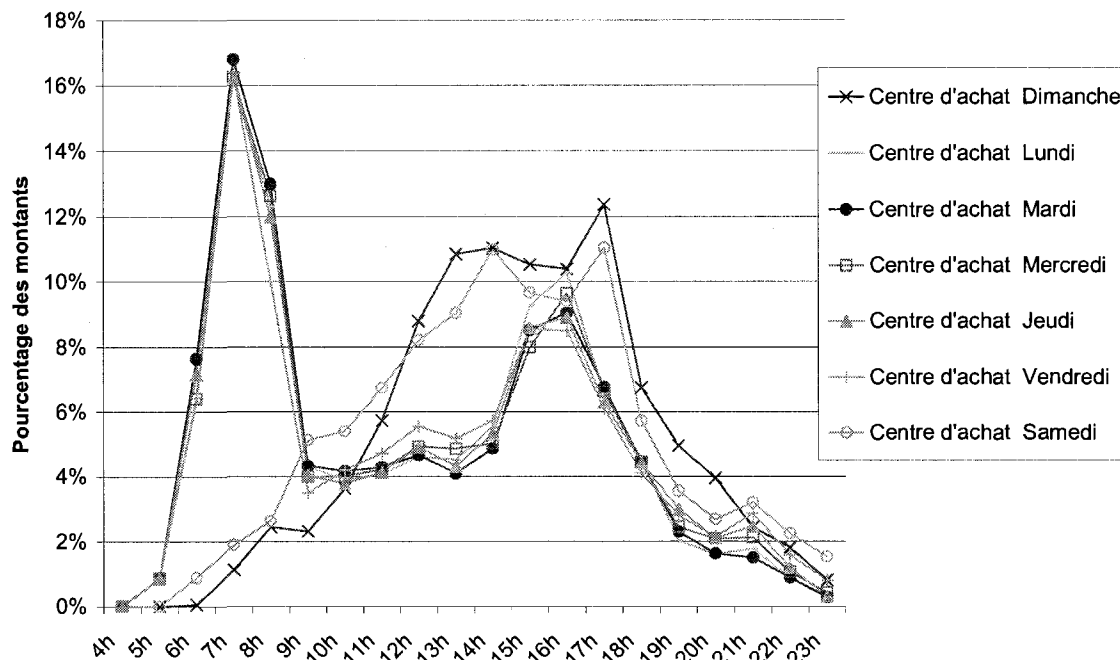


Figure 3.35. Profil des montants aux arrêts proches des centres d'achat par heure et par jour de la semaine

En observant la Figure 3.35, on se rend compte que, comme pour les écoles, les comportements à proximité des centres commerciaux (centres d'achats) sont très semblables du lundi au vendredi. Ici encore le vendredi se différencie des autres jours avec une augmentation de l'utilisation pendant la période du midi. On voit aussi que le pic d'activité maximum des arrêts situés à proximité d'un centre d'achat correspond à l'heure de pointe du matin, les centres commerciaux sont donc un point de départ du matin, cela est principalement dû au fait que certains stationnements de centres commerciaux sont utilisés comme parc-o-bus. On constate aussi que le samedi et le dimanche, on a un pic d'activité entre 17h et 18h qui correspond au retour de

magasinage du fait de la fermeture des commerces à 17h (de même les jeudis et vendredi avec la fermeture des magasins à 21h).

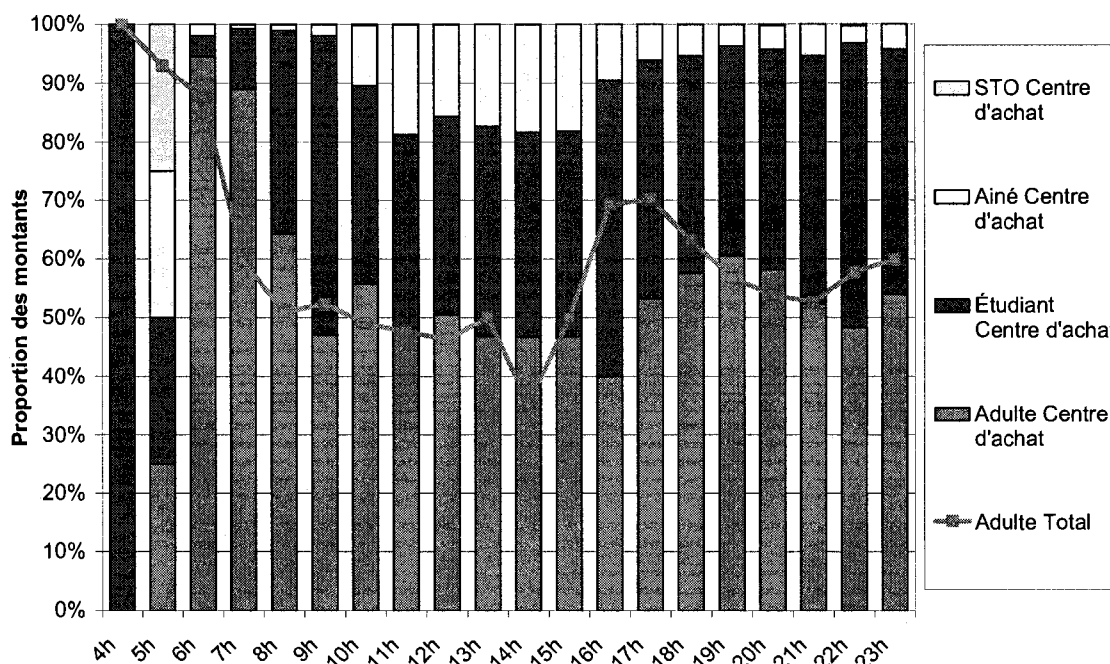


Figure 3.36. Proportion des classes d'utilisateurs montant aux arrêts proches des centres commerciaux pour tous les jeudis de septembre 2005 par heure de la journée

La figure précédente (Figure 3.36) est construite de la même manière que la Figure 3.34. Si on compare la proportion de chaque classe d'utilisateurs montant aux arrêts proches des centres d'achat (Figure 3.36) à la proportion moyenne (Figure 2.17), on se rend compte qu'entre 10h du matin et 18h les aînés sont nombreux en proportion à monter à des arrêts proches des centres commerciaux. Cela pourrait s'expliquer par le fait que les aînés ont le temps la journée pour aller se promener dans les centres d'achats. Entre 16h et 19h, il y a beaucoup plus d'étudiants en proportion aux arrêts proches des centres commerciaux. Cela peut, par exemple, correspondre aux étudiant(e)s qui vont magasiner après les cours et qui rentrent chez eux avant le souper.

3.4.3. Caractérisation des lieux d'activités

On a vu dans la partie 3.2.1.3 que l'on pouvait classer la plupart des arrêt-ligne-direction dans un groupe prépondérant. Les arrêt-lignes-direction qui varient beaucoup et qui n'ont pas de groupe prépondérant sont rangés dans le groupe «Aucun». Cela permet de diminuer le bruit dans les données en identifiant les arrêts les plus imprévisibles. On agrège ensuite le tableau précédent par groupe prépondérant et on compte le nombre d'arrêt-ligne-direction dans chaque groupe prépondérant en relation de proximité avec les générateurs de déplacements. On obtient donc la répartition suivante des arrêts-ligne-direction (Tableau 3.14) :

Tableau 3.14. Groupes prépondérants et générateurs de déplacements

Groupe prépondérant	Proche parc-o-bus	Proche École	Inclus dans centre ville	Proche centre d'achat	Proche hôpital	Nœud
AUCUN	145	247	203	107	95	6
Groupe régulier	63	153	98	70	39	62
Groupe pic 6h	9	13	4	16	4	1
Groupe pic 7h	66	58	49	30	27	5
Groupe pic 8h	26	63	40	27	24	0
Groupe pic Milieu	40	89	51	40	39	1
Groupe pic 15h	50	78	23	23	15	0
Groupe pic 16h	43	110	212	26	17	3
Groupe pic 17h	31	42	39	16	12	0
Groupe pic Soir	53	56	50	30	8	3
Total	526	909	769	385	280	81

Si on lit le Tableau 3.14 par colonne, on obtient une caractérisation des générateurs de déplacements en fonction de leur proximité aux groupes d'arrêt-ligne-direction. Le

Tableau 3.15 présente le pourcentage d'arrêt-ligne-direction de chaque groupe prépondérant aux alentours des lieux d'activités.

Tableau 3.15. Pourcentage des différents groupes prépondérants proches des générateurs de déplacements

Groupe prépondérant	Proche parcobus	Proche École	Inclus dans centre ville	Proche centre d'achat	Proche hôpital	Nœud
AUCUN	27,57%	27,17%	26,40%	27,79%	33,93%	7,41%
Groupe régulier	11,98%	16,83%	12,74%	18,18%	13,93%	76,54%
Groupe pic 6h	1,71%	1,43%	0,52%	4,16%	1,43%	1,23%
Groupe pic 7h	12,55%	6,38%	6,37%	7,79%	9,64%	6,17%
Groupe pic 8h	4,94%	6,93%	5,20%	7,01%	8,57%	0,00%
Groupe pic Milieu	7,60%	9,79%	6,63%	10,39%	13,93%	1,23%
Groupe pic 15h	9,51%	8,58%	2,99%	5,97%	5,36%	0,00%
Groupe pic 16h	8,17%	12,10%	27,57%	6,75%	6,07%	3,70%
Groupe pic 17h	5,89%	4,62%	5,07%	4,16%	4,29%	0,00%
Groupe pic Soir	10,08%	6,16%	6,50%	7,79%	2,86%	3,70%
Total	100%	100%	100%	100%	100%	100%

Ainsi, on peut voir qu'à proximité de pratiquement tous les générateurs de déplacements, on trouve environ 30% d'arrêt-ligne-direction dont le comportement est imprévisible. Ce pourcentage correspond au pourcentage moyen des arrêt-ligne-direction qui ne sont dans aucun groupe prépondérant.

Les arrêts proches des parc-o-bus ont tendance à être surtout utilisés :

- entre 7h et 8h
- régulièrement toute la journée

- le soir

Les arrêt-ligne-direction proches des écoles ont tendance à être surtout utilisés :

- régulièrement toute la journée
- dans le milieu de la journée
- entre 15h et 17h

Les arrêt-ligne-direction du centre ville ont tendance à être surtout utilisés :

- entre 16h et 17h, c'est la période d'influence maximale l'après-midi, la période de pointe de l'après-midi

Les arrêt-ligne-direction proches des centres commerciaux ont tendance à être surtout utilisés :

- régulièrement
- dans le milieu de la journée
- le soir

Les arrêt-ligne-direction proches des hôpitaux ont tendance à être surtout utilisés :

- régulièrement dans la journée
- dans le milieu de la journée
- pas du tout le soir

Les nœuds du réseau engendrent des arrêt-ligne-direction qui sont utilisés toute la journée (plus de 75% des nœuds sont dans le noyau du groupe Régulier).

On voit donc que les lieux d'activités génèrent une utilisation particulière autour d'eux. Chaque lieu impose une marque caractéristique sur le réseau de transport en commun. On peut cependant penser que cette marque varie au cours de la semaine. Par exemple, il

doit y avoir une différence pour le centre ville entre la semaine (du lundi au vendredi) et la fin de semaine.

3.4.3.1. Variabilité de l'utilisation des lieux d'activités en fonction du jour de la semaine

Nous allons étudier pour chaque type de générateurs de déplacements la variabilité de son influence sur le réseau de transport en fonction des jours de la semaine.

On a vu dans la partie 3.2.1.3 qu'il était possible de définir un groupe prépondérant (avec un taux caractéristique d'appartenance de plus de 50%) pour chaque arrêt-ligne-direction pour chaque jour de la semaine. La variabilité des comportements des arrêt-ligne-direction est plus faible un jour de la semaine donné que globalement sur la semaine. Ainsi, on a vu que pour les jours du lundi au vendredi, plus de 85% des arrêt-ligne-direction appartiennent au noyau d'un groupe (78% si on considère tous les jours) et que l'utilisation du réseau le week-end est beaucoup plus irrégulière.

Par exemple, on construit le même tableau que le Tableau 3.14 mais en considérant seulement les variations les lundis (voir Tableau 3.16).

Tableau 3.16. Groupes prépondérants du lundi et générateurs de déplacements

Gp prépondérant lundi	Proche parcobus	Proche Ecole	Inclus dans centre ville	Proche centre d'achat	Proche hopitaux	Nœud
AUCUN	77	129	110	64	56	2
Groupe régulier	72	165	96	79	45	61
Groupe pic 6h	12	17	8	15	5	1
Groupe pic 7h	71	67	52	30	29	6
Groupe pic 8h	28	61	36	25	25	0
Groupe pic Milieu	44	92	51	40	39	1
Groupe pic 15h	52	76	43	25	12	0
Groupe pic 16h	42	118	220	28	24	4
Groupe pic 17h	31	42	33	18	9	0
Groupe pic Soir	54	62	57	34	13	6
Total	483	829	706	358	257	81

En pourcentage par colonne, le tableau précédent devient :

Tableau 3.17. Pourcentage des différents groupes prépondérants du lundi proches des générateurs de déplacements

Gp prépondérant lundi	Proche parcobus	Proche École	Inclus dans centre ville	Proche centre d'achat	Proche hopitaux	Nœud
AUCUN	15,94%	15,56%	15,58%	17,88%	21,79%	2,47%
Groupe régulier	14,91%	19,90%	13,60%	22,07%	17,51%	75,31%
Groupe pic 6h	2,48%	2,05%	1,13%	4,19%	1,95%	1,23%
Groupe pic 7h	14,70%	8,08%	7,37%	8,38%	11,28%	7,41%
Groupe pic 8h	5,80%	7,36%	5,10%	6,98%	9,73%	0,00%
Groupe pic Milieu	9,11%	11,10%	7,22%	11,17%	15,18%	1,23%
Groupe pic 15h	10,77%	9,17%	6,09%	6,98%	4,67%	0,00%
Groupe pic 16h	8,70%	14,23%	31,16%	7,82%	9,34%	4,94%
Groupe pic 17h	6,42%	5,07%	4,67%	5,03%	3,50%	0,00%
Groupe pic Soir	11,18%	7,48%	8,07%	9,50%	5,06%	7,41%

On voit dans le Tableau 3.17 que le nombre d'arrêt-ligne-direction très variables (ceux qui ne sont dans AUCUN noyau) diminue beaucoup si on considère seulement les lundis. C'est normal, on a vu que l'utilisation est plus régulière d'un lundi sur l'autre qu'en général.

On peut donc calculer pour chaque générateur de déplacements la répartition des groupes prépondérants d'arrêt-ligne-direction proches pour tous les jours de la semaine. Prenons les exemples des centres commerciaux et des écoles.

3.4.3.2. Étude de la variabilité hebdomadaire des arrêts proches des écoles :

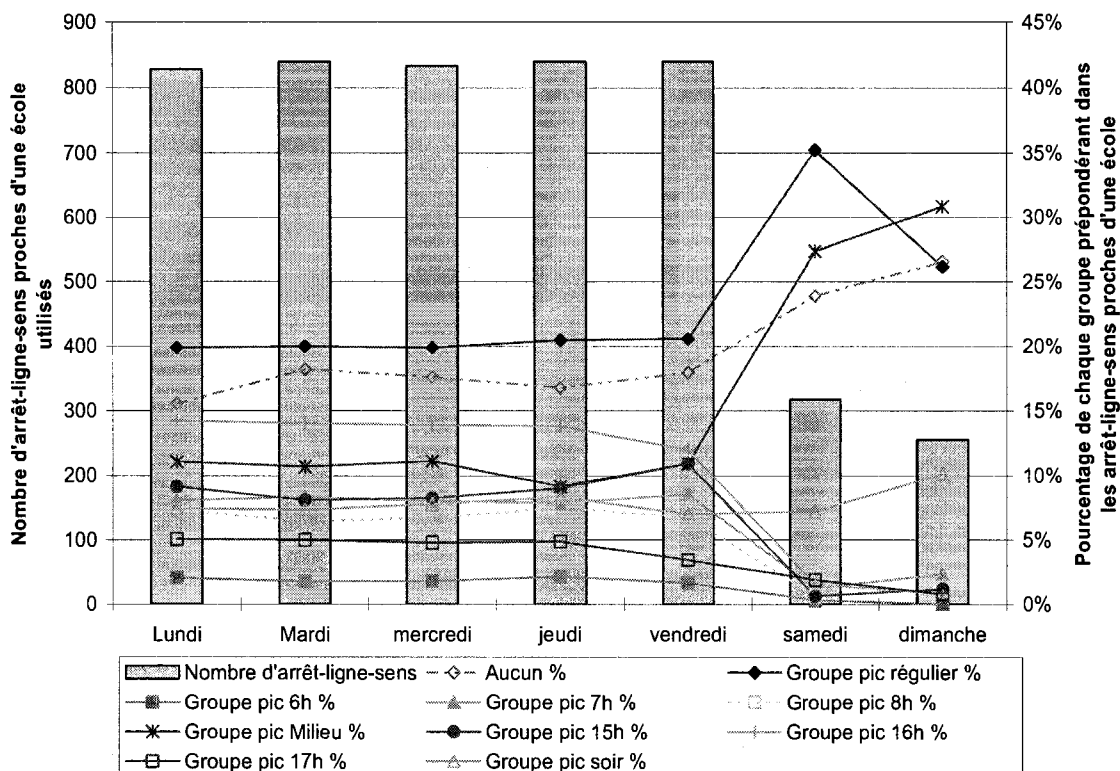


Figure 3.37. Variabilité de la proportion des arrêt-ligne-direction proches des écoles dans les groupes prépondérants pour les différents jours de la semaine.

Les comportements des arrêt-ligne-direction proche des écoles sont extrêmement réguliers du lundi au vendredi. On a tous les jours quasiment le même nombre d'arrêts de chaque type (voir Figure 3.37). Seul le vendredi est un peu différent, la proportion des groupes qui ont un pic d'activité après 16h a tendance à diminuer, alors que la proportion des groupes pic milieu et pic 15h augmente faiblement. Cela veut dire que les arrêts proches des écoles sont utilisés plus tôt l'après-midi le vendredi que les autres jours de la semaine. Le week-end, les écoles sont fermées, les écoles n'influencent donc pas vraiment le réseau, on se rapproche de la répartition moyenne dans les différents noyaux.

Mais, cependant, on note beaucoup d'activité le dimanche soir. La proportion des classes d'utilisateurs qui montent le dimanche aux arrêts proches des écoles et qui ont comme groupe prépondérant le groupe pic Soir est la même que la proportion moyenne des classes d'utilisateurs. Du coup, rien ne nous a permis d'expliquer ce pic d'activité.

3.4.3.3. Étude de la variabilité hebdomadaire des arrêts proches des centres d'achat :

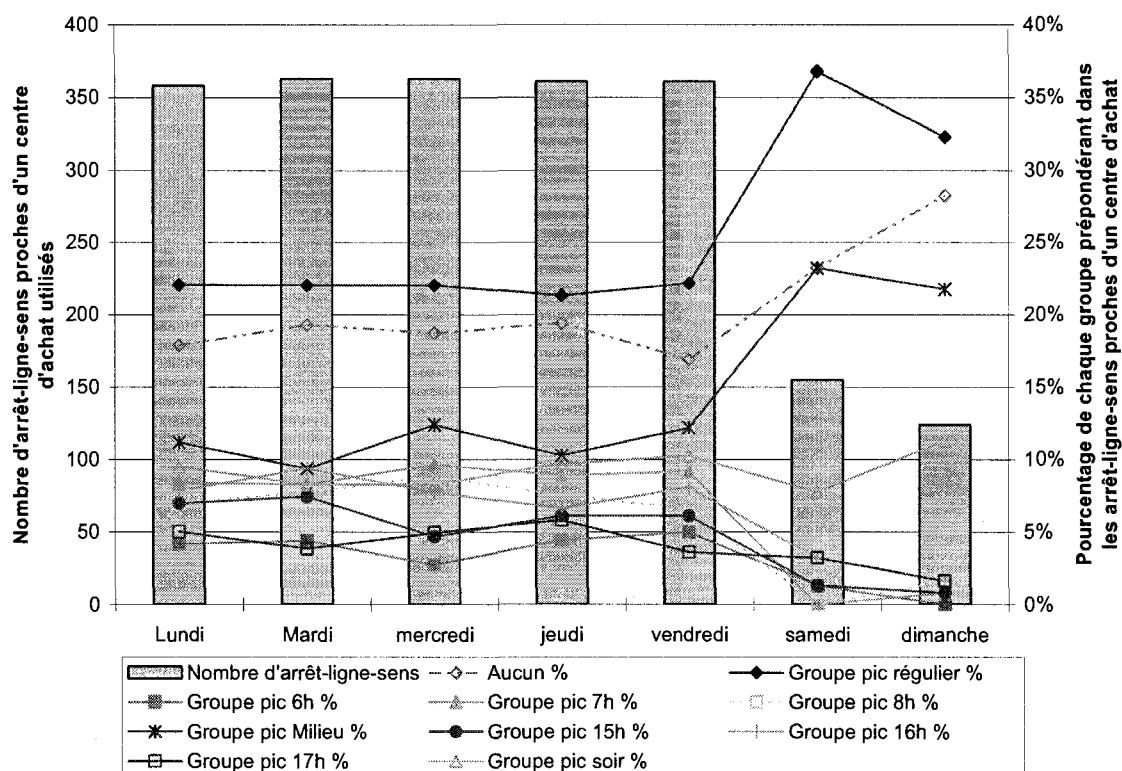


Figure 3.38. Variabilité de la proportion des arrêt-ligne-direction proches des centres commerciaux dans les groupes prépondérants pour les différents jours de la semaine

La Figure 3.38 se lit ainsi :

- les courbes montrent que 22% des arrêt-ligne-direction proche d'un centre commercial appartiennent au noyau du groupe régulier les lundis.

- l'histogramme montre le nombre d'arrêt-ligne-direction utilisés proches d'un centre d'achat pour chaque jour de la semaine.

On voit déjà que le nombre d'arrêt-ligne-direction utilisés près d'un centre commercial varie peu du lundi au vendredi (environ 360). Le week-end, il chute fortement pour tomber à 160 le samedi et moins de 130 le dimanche. La part des arrêt-ligne-direction qui varient beaucoup passe de 18% la semaine à 23% et 28% le week-end. Pourtant, on aurait pu penser que les arrêts proches des centres commerciaux induisaient, le week-end, une activité assez régulière (d'un samedi sur l'autre par exemple) sur le réseau de transport en commun. Or, on se retrouve avec une proportion d'arrêts irréguliers proche de la moyenne.

De même, la part des deux principaux groupes est assez constante pendant la semaine, 22% pour le groupe Régulier et 18% pour celui des arrêts imprévisibles. Le week-end les arrêts proches de centre commerciaux sont surtout utilisés régulièrement toute la journée ou au milieu de la journée ou encore de manière irrégulière.

Il est difficile d'analyser plus finement car un arrêt donné est souvent situé près d'un ensemble de générateurs de déplacements. L'influence des centres commerciaux est donc un peu « polluée ». Ainsi, on s'attend à ce que plus d'arrêt-ligne-direction soient utilisés le soir les jeudis et vendredis que les lundis, mardis et mercredis. En effet, la plupart des magasins ferment à 21h le jeudi et le vendredi. Or le lundi, on a quasiment autant d'arrêt-ligne-direction dans le groupe pic Soir que le jeudi ou le vendredi.

On voit donc que les différents lieux d'activités identifiés engendrent une activité caractéristique sur le réseau de transport en commun. Un générateur de déplacements imprime une marque particulière sur les comportements des arrêts de bus qui lui sont proches. Cette empreinte varie assez peu pendant la semaine (du lundi au vendredi) mais est très différente, beaucoup moins régulière et plus imprévisible les fins de semaine.

CONCLUSION

Ce mémoire montre les potentialités de l'analyse des données de cartes à puce pour comprendre l'impact des lieux d'activités d'un territoire sur l'utilisation d'un réseau de transport collectif desservant ce territoire. Pour arriver à cette conclusion, nous avons cherché à caractériser l'influence des principaux générateurs de déplacements sur les comportements des arrêts situés à proximité.

Bilan

Après avoir fait une revue de littérature dans les domaines pertinents pour cette étude, nous avons présenté le contexte de la STO et les tendances principales du réseau de bus. Puis, nous avons expliqué la méthodologie utilisée pour parvenir aux objectifs. Ensuite, nous avons construit un SIG de l'Outaouais. Pour ce faire, nous avons, tout d'abord, identifié et localisé les principaux générateurs de déplacements du territoire ; ce sont le centre ville, les écoles, les stationnements incitatifs, les hôpitaux, les centres commerciaux et les nœuds du réseau.

Ensuite, nous avons étudié les comportements des arrêts de bus du réseau en segmentant la population d'arrêts en un petit nombre de groupes ayant des comportements types. Ensuite, nous avons discuté le choix de ce petit nombre de groupes pour arriver à un nombre adéquat de groupes du point de vue de la compréhension du réseau. On a obtenu une segmentation des arrêts en fonction d'un pic d'utilisation à une période donnée de la journée.

Puis, en étudiant la variabilité de l'appartenance des arrêts à ces groupes de comportements types au cours du temps, nous avons réussi à diminuer le bruit dans les

données en séparant les arrêts très variables des autres arrêts. Cette étude de variabilité a aussi permis de montrer la grande répétitivité des comportements spatio-temporelles pendant les jours de la semaine (du lundi au vendredi), ce patron de comportements est encore plus répétitif pour un jour de la semaine. Autrement dit, si on connaît les déplacements des usagers pour un jour donné (par exemple un lundi), on peut prévoir les déplacements des usagers pour un autre lundi sans trop se tromper.

Ensuite, nous avons analysé les comportements des usagers montant sur des arrêts appartenant à un groupe donné. On a pu voir que :

- les arrêts qui ont un pic d'utilisation pendant l'heure creuse du milieu de la journée sont en majorité utilisés par des étudiants tandis que les arrêts des heures de pointe sont surtout utilisés par des adultes.
- Plus les usagers habitent loin du centre ville, plus ils utilisent les arrêts qui ont un pic d'activité tôt le matin pour faire des déplacements plus longs.

En analysant les relations de proximité entre les différents éléments du SIG, nous avons pu voir que chaque générateur de déplacements imprime une marque particulière sur le réseau de transport en commun tant au niveau de la répartition des montants dans la journée qu'au niveau de la classe des montants. Ceci permet de caractériser l'influence directe de ces lieux d'activités sur l'utilisation des arrêts de bus. On a aussi pu observer les variations hebdomadaires des empreintes des générateurs de déplacements et voir qu'elles sont relativement homogènes du lundi au vendredi, mais que la variabilité de ces empreintes augmente beaucoup les fins de semaine pour quasiment tous les types de générateurs de déplacements identifiés.

Contributions

Cette étude a permis de proposer une méthode pour répondre à une question fondamentale lorsque l'on fait des segmentations de population en datamining, la question du choix du nombre de groupes. Pour ce faire, on a travaillé sur la façon de

représenter les données en regroupant les zones où l'on voulait moins de définition. Il nous a fallu faire un compromis entre niveau de précision et significativité des groupes.

Ce mémoire a aussi été l'occasion de développer des méthodes de traitement et de représentation des données. Ainsi, nous avons regroupé les arrêts les plus réguliers dans des noyaux qui contiennent donc les éléments les plus significatifs des différents groupes. De plus, nous avons construit un SIG de l'Outaouais qui nous a permis de relier les analyses de datamining sur les arrêts à leur environnement géographique réel.

Finalement, cette étude a permis d'avoir une meilleure connaissance des générateurs de déplacements de l'Outaouais en analysant l'empreinte des principaux lieux d'activités de la région sur les comportements des utilisateurs de cartes à puce.

Perspectives

Le SIG que l'on a créé dans cette étude pourrait être plus précis en identifiant plus de lieux d'activités comme par exemple les zones industrielles principales de l'Outaouais ou les écoles primaires. Il serait intéressant de profiter des contributions de ce mémoire pour étudier de nouvelles politiques tarifaires reliées aux générateurs de déplacements. Il serait également intéressant d'analyser l'offre de transport collectif de façon détaillée (lignes, voyages, véhicules) en étudiant les charges à bord des autobus et les clientèles qui s'y trouvent. On pourrait aussi imaginer utiliser la connaissances des variations spatiotemporelles de l'utilisation du bus pour alimenter un système publicitaire variable en fonction du temps et de arrêts desservis pour que l'offre publicitaire coïncide le plus possible avec les types d'usagers par exemple.

On pourrait affiner et compléter les analyses en croisant l'étude des données de cartes à puce avec les résultats des enquêtes origine-destination de l'Outaouais. En effet, les données récoltées lors de ces enquêtes sont plus complètes quant aux raisons de

déplacements et aux caractéristiques socio-démographiques des usagers par exemple. L'analyse en parallèle de ces deux sources de données peut servir à valider les observations et les études faites à partir d'une seule de ces deux sources.

BIBLIOGRAPHIE

AXHAUSEN, K.W., ZIMMERMANN, A., SCHÖNFELDER, S., RINDSFÜSER, G., HAUPT, T. (2002). Observing the rhythms of daily life: a six-week travel diary. *Transportation*, 29 (2), 95–124.

BADDOE, D., MILLER, E. (2000). Transportation-land use interaction: empirical findings in North America, and their implications for modeling. *Transportation Research Part, D5* (4), 235–263.

BAGCHI, M., WHITE, P.R. (2004). What role for smart-card data from a bus system? *Municipal Engineer*, 157, 39–46.

BAGCHI, M., WHITE, P.R. (2005). The potential of public transport smart card data. *Transport Policy*, 12, 464–474.

BALMER, A., FRENCH, J., ECK, R., LEGG, J. (2000) Trip Generation Rates of Consolidated Schools. *Institute of Transportation Engineers Journal*, Vol. 70(8), 30-34.

BERRY, M., LINOFF, G. (1997). *Data Mining Techniques: for Marketing, Sales, and Customer Support*. Wiley, New York.

BLYTHE, P. (1998). Integrated ticketing smart cards in transport. *IEE Colloquium: Using ITS in Public Transport and in Emergency Services*, 1–21.

BLYTHE, P. (2004). Improving public transport ticketing through smart cards. *Proceedings of the Institution of Civil Engineers. Municipal Engineer* 157, 47–54 (March Issue ME1).

BONNEAU, W. AND EDITORS. (2002). The role of smart cards in mass transit systems, *Card Technology Today*, June 2002, 10.

CERVERO, R., KOCKELMAN, K. (1997). Travel demand and the 3D's: density, diversity, and design. *Transportation Research, D 2 (3)*, 199-219.

CERVERO, R., LANDIS, J. (1997). Twenty years of the Bay Area rapid transit system: land-use and development impacts. *Transportation Research, A 31 (4)*, 309-333.

CHAPLEAU, R. (1995). Symphonie d'usages des grandes enquêtes Origine-Destination, en totalement désagrégé majeur, opus Montréal 87 et 93. Eightieth "Entretiens du Centre Jacques-Cartier". Lyon.

CLARKE, R. (2001). Person location and person tracking: technologies, risks and policy implications. *Information Technology and People, 14 (2)*, 206-231.

CNIL –Commission nationale de l'informatique et des libertés. (2003). Recommandation relative à la collecte et au traitement d'informations nominatives par les sociétés de transports collectifs dans le cadre d'applications billettiques, CNIL, Délibération N° 03-038.

DATTA, T.K., DATTA, S., NANNAPANENI, P. (1998). Trip generation models for multiuse highway commercial developments. *Institute of Transportation Engineers Journal, Vol. 68(2)*, 24-30.

DOHERTY, S.T., MILLER, E.J. (2000). A computerized household activity scheduling survey. *Transportation, 27 (1)*, 75-97.

FAYYAD, U.M., PIATETSKY-SHAPIO, G., SMYTH, P., UTHURUSAMY, R. (1996). *Advances in Knowledge Discovery and Data Mining*. AAAI Press, The MIT Press, Cambridge, MA.

GÄRLING, T., AXHAUSEN, K.W. (2003). Introduction: habitual travel choice. *Transportation*, 30 (1), 1–11.

HOFFMAN, M., O'MAHONY, M. (2005). Transfer Journey Identification and Analyses from Electronic Fare Collection Data. *Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems*, Vienna, Austria, September 13–16, 2005, 825–830.

HUNT, J.D., MCMILLAN, J.D.P., ABRAHAM, J.E. (1994). Stated preference investigation of influences on attractiveness of residential locations. *Transportation Research Record* 1466, 79–87.

JOHNSON, K.L., HAMMOND, M.I. (2001). Trip generation characteristics for convenience stores. *Institute of Transportation Engineers Journal*, Vol.71(8), 26–30.

Jones, P., Clark, M. (1988). The significance and measurement of variability in travel behaviour. *Transportation*, 15(1), 65–87.

JUN, M., GOULIAS, K. (1997). A dynamic analysis of person and household activity and travel patterns using data from the first two waves in the Puget sound transportation panel. *Transportation*, (24), 309–331.

KITAMURA, R. (1988). An evaluation of activity-based travel analysis. *Transportation*, 15 (1), 9–34.

KITAMURA, R., YAMAMOTO, T., SUSILO, Y.O., AXHAUSEN, K.W. (2006). How routine is a routine? An analysis of the day-to-day variability in prism vertex location. *Transportation Research, Part A* (40), 259–279.

KNIGHT, R.L., TRYGG, L.L. (1977). Land-use impacts of rapid transit systems: implications of recent experience. *Final Report Prepared for the US Department of Transportation*.

KOPERSKI, K. AND HAN, J. (1995). Discovery of spatial association rules in geographic information databases. *4th Int. Symp. on Large Spatial Databases (SSD '95)*, Portland, ME, 47–66.

KOPERSKI, K., ADHIKARY, J., AND HAN, J. (1996). Knowledge discovery in spatial databases: Progress and challenges. *SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. Technical Report 96-08*, University of British Columbia, Vancouver, Canada.

LEE.M.S, MCNALLY M.G. (2003). On the structure of weekly activity/travel patterns. *Transportation Research, Part A* 37, 823-839.

LOCKWOOD, A., SRINIVASAN, S., BHAT, C.R. (2005). An exploratory analysis of weekend activity patterns in the San Francisco Bay area. *Transport. Res. Rec.* 1926, 70–78.

MEADOWCROFT, P. (2005). Hong Kong raises the bar in smart card innovation. *Card Technology Today*, 17 (1), 12–13.

MORENCY, C., TRÉPANIÉ, M., AGARD, B. (2006). Analysing the variability of transit users behaviour with smart card data. *The Ninth International IEEE Conference on Intelligent Transportation Systems*, Toronto, Canada, September.

MORENCY, C., TRÉPANIÉ, M., AGARD, B. (2007). Measuring transit use variability with smart-card data. *Transport Policy*, Volume 14, Issue 3, 193-203.

MURAKAMI, E., WAGNER, D.P., NEUMEISTER, D.M. (2000). Using Global Positioning Systems and Personal Digital Assistants for Personal Travel Surveys in the United States. *Transport Surveys: Raising the Standard, Proceedings of an International Conference on Transport Survey Quality and Innovation, TRB Transportation Research Circular EC008*, iii-B/1 – iii-B/21.

RAKOTOMALALA, R. (2005). TANAGRA : un logiciel gratuit pour l'enseignement et la recherche. *Actes de EGC'2005, RNTI-E-3, vol. 2*, 697-702.

SCHLICH, R., AXHAUSEN, K.W. (2003). Habitual travel behaviour: evidence from a six-week travel diary. *Transportation*, (30), 13–36.

SECRÉTARIAT À L'ENQUÊTE ORIGINE-DESTINATION. (2005). Enquête Origine-Destination 2003. Mobilité des personnes dans la région de Montréal. Agence métropolitaine de transport. 176 pages.

SHELFER, K.M., PROCACCINO, J.D. (2002). Smart card evolution. *Communications of the ACM*, 45 (7), 83–88.

TRÉPANIÉ, M., CHAPLEAU, R., ALLARD, B., MORENCY, C. (2003). Trip generator relocation impact analysis methodology based on household surveys. *Institute of Transportation Engineers Journal*, Vol. 73(10).

TRÉPANIÉ, M., CHAPLEAU, R. (2001). Analyse orientée-objet et totalement désagrégée des données d'enquêtes ménages origine-destination. *Revue Canadienne de Génie Civil, Ottawa* 28 (1), 48–58.

TRÉPANIÉ, M., CHAPLEAU, R. (2006). Destination estimation from public transport smartcard data. *The 12th IFAC Symposium on Information Control Problems in Manufacturing (INCOM)*, Saint-Étienne, France.

VIVIAN G.M. (2006). Trip generation characteristics of free-standing Discount Superstores. *Institute of Transportation Engineers Journal, Vol. 76(8)*, 30-37.

WEEKS J. (2004). The role of spatial analysis in demographic research. *Spatially integrated social science*. Goodchild MF and Janelle DG (eds.). Oxford University Press.

WESTPHAL, C., BLAXTON, T. (1998). *Data Mining Solutions*. Wiley, New York.

WORKMAN, S.L., BROD, D. (1997). Measuring the neighborhood benefitts of rail transit accessibility. Presented at 76th Annual Meeting of the Transportation Research Board, Washington DC.

Sites web

ALEXANDRE K. ET EVGENIY S. *Let's describe the world*. Consulté le 15 septembre 2007, tiré de <http://wikimapia.org>.

COMMISSION SCOLAIRE DES PORTAGES-DE-L'OUTAOUAIS. *Commission Scolaire des Portages-de-l'Outaouais*. Consulté le 15 septembre 2007, tiré de <http://www.cspo.qc.ca>

DIRECTION DES SYSTÈMES D'INFORMATION (DSI), SECTION DU DÉVELOPPEMENT GÉOMATIQUE. *Atlas de Gatineau*. Consulté le 5 septembre 2007, tiré de <http://www.ville.gatineau.qc.ca/website/AtlasInternet/asp.htm>

GENDRON A. *Cégep de l'Outaouais*. Consulté le 10 septembre 2007, tiré de <http://www.cegepoutaouais.qc.ca/>

GOOGLE. *Google Maps*. Consulté le 16 septembre 2007, tiré de <http://maps.google.ca>.